PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [6] :
C07H 21/02, 21/04, C12N 1/00, 5/06, 15/00, 15/09, 15/11, C12P 19/34, 21/02, 21/06, C12Q 1/00, 1/70, G01N 33/53

A1

(11) International Publication Number: WO 96/38462

(43) International Publication Date: 5 December 1996 (05.12.96)

(21) International Application Number: PCT/US96/08295

(22) International Filing Date: 31 May 1996 (31.05.96)

(30) Priority Data:

08/460,505	2 June 1995 (02.06.95)	US
08/588,258	12 January 1996 (12.01.96)	US

(71) Applicant: MASSACHUSETTS INSTITUTE OF TECHNOLOGY [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02142-1324 (US).

(72) Inventors: HORVITZ, H., Robert; 54 Maple Street, Auburndale, MA 02166 (US). KOELLE, Michael; 90 Morrison Avenue, Somerville, MA 02144 (US).

(74) Agent: CLARK, Paul, T.; Fish & Richardson. P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: REGULATORS OF G-PROTEIN SIGNALLING

(57) Abstract

Disclosed is substantially pure DNA encoding a *C. elegans* EGL-10 polypeptide; substantially pure EGL-10 polypeptide; methods of obtaining *rgs* encoding DNA and RGS polypeptides; and methods of using the *rgs* DNA and RGS polypeptides to regulate G-protein signalling.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

REGULATORS OF G-PROTEIN SIGNALLING

Background of the Invention

The invention relates to regulators of heterotrimeric G-protein mediated events and uses thereof to mediate cell signalling and membrane trafficking.

The heterotrimeric guanine nucleotide binding proteins (G proteins) are intracellular proteins best known for their role as transducers of binding by extracellular ligands to seven transmembrane receptors (7-TMRs) located on the cell surface. Individual 7-TMRs have been identified for many small neurotransmitters (e.g. adrenaline, noradrenaline, dopamine, serotonin, histamine, acetylcholine, GABA, glutamate, and adenosine), for a variety of neuropeptides and hormones (e.g. opioids, tachykinins, bradykinins, releasing hormones, vasoactive intestinal peptide, neuropeptide Y, thyrotrophic hormone, leutenizing hormone, follicle-stimulating hormone, adrenocorticotropic hormone, cholecystokinin, gastrin, glucagon, somatostatin, endothelin, vasopressin and oxytocin) as well as for chemoattractant chemokines (C5a, interleukin-8, platelet-activating factor and the N-formyl peptides) that are involved in immune function. In addition, the odorant receptors present on vertebrate olfactory cells are 7-TMRs, as are rhodopsins, the proteins that transduce visual signals.

Ligand binding to 7-TMRs produces activation of one or more heterotrimeric G-proteins. A few proteins with structures that are dissimilar to the 7-TMRs have also been shown to activate heterotrimeric G-proteins. These include the amyloid precursor protein, the terminal

complement complex, the insulin-like growth factor/mannose 6-phosphate receptor and the ubiquitous brain protein GAP-43. Dysregulation of G-protein coupled pathways is associated with a wide variety of diseases, including diabetes, hyperplasia, psychiatric disorders, cardiovascular disease, and possibly Alzheimer's disease. Accordingly, the 7-TMRs are targets for a large number of therapeutic drugs: for example, the β-adrenergic blockers used to treat hypertension target 7-TMRS.

Unactivated heterotrimeric G-proteins are complexes comprised of three subunits, Gα, Gβ and Gγ. The subunits are encoded by three families of genes: in mammals there are at least 15 Gα, 5 Gβ and 7 Gγ genes. Additional diversity is generated by alternate splicing. Where it has been studied, a similar multiplicity of G-proteins has been found in invertebrate animals. Mutations within Gα subunit genes is involved in the pathophysiology of several human diseases: mutations of Gα that activate Gs or Gi2 are observed in some endocrin tumors and are responsible for McCune-Albright syndrome, whereas loss-of-function mutations of Gαs are found in Albright hereditary osteodystrophy.

The Gα subunits have binding sites for a guanine nucleotide and intrinsic GTPase activity. This structure and associated mechanism are shared with the monomeric GTP-binding proteins of the ras superfamily. Prior to activation the complex contains bound GDP: GαGDP$\beta\gamma$. Activation involves the catalyzed release of GDP followed by binding of GTP and concurrent dissociation of the complex into two signalling complexes: GαGTP and $\beta\gamma$. Signalling through GαGTP, the more thoroughly characterized pathway, is terminated by GTP hydrolysis to GDP. GαGDP then reassociates with $\beta\gamma$ to reform the inactive, heterotrimeric complex.

The mammalian G-proteins are divided into four subtypes: Gs, Gi/Go, Gq and G12. This typing is based on the effect of activated G-proteins on enzymes that generate second messengers and on their sensitivity to cholera and pertussis toxin. These divisions also appear to be evolutionarily ancient: there are comparable subtypes in invertebrate animals. Members of two subtypes of G-proteins control the activity of adenylyl cyclases (ACs). Activated Gs proteins increase the activity of ACs whereas activated Gi proteins (but not Go) inhibit these enzymes. Gs proteins are also uniquely activated by cholera toxin. ACs are the enzymes responsible for the synthesis of cyclic adenosine monophosphate (cAMP). cAMP is a diffusible second messenger that acts through cAMP-dependent protein kinases (PKAs) to phosphorylate a large number of target proteins. Members of two subtypes, all Gi/Go proteins and the Gq proteins, increase the activity of inositol phospholipid-specific phospholipases (IP-PLCs). The activity of the subtypes are distinguishable: activation of Gi and Go are blocked by pertussis toxin whereas Gq is resistant to this compound. IP-PLCs release two diffusible second messengers, inositol triphosphate (IP_3) and diacylglycerol (DAG). IP_3 modulates intracellular Ca^{2+} concentration whereas DAG activates protein kinase Cs (PKCs) to phosphorylate many target proteins. The second messenger cascades allow signals generated by G-protein activation to have global effects on cellular physiology.

Activation of G proteins frequently modulate ion conductance through plasma membrane ion channels. Although in some cases these effects are indirect, as a result of changes in second messengers, G-proteins can also couple directly to ion channels. This phenomenon is known as membrane delimited modulation. The opening of

inwardly rectifying K channels by activated Gi/Go and of N and L type Ca channels by Gi/Go and Gq are commonly observed forms of membrane delimited modulation.

Heterotrimeric G proteins appear to have other cellular roles, in addition to transducing the binding of extracellular ligands. Analysis of the intracellular localization of the various G-protein subunits combined with pharmacological studies suggest, for example, that G proteins are involved in intracellular membrane trafficking. Indeed, some workers hypothesize that G proteins evolved to control membrane trafficking and that their role in transducing extracellular signals evolved later. Studies implicate heterotrimeric G-proteins in the formation of vesicles from the trans-Golgi network, in transcytosis in polarized epithelial cells and in the control of secretion in many cells, including several model systems relevant to human disease: mast cells, chromaffin cells of the adrenal medulla and human airway epithelial cells. Nonetheless, the G-protein subunits involved in membrane trafficking and secretion have yet to be definitively established and the mechanisms by which they are activated and control membrane trafficking remains largely unknown.

Caenorhabditis elegans (reviewed in Wood, et al. (1988) *The Nematode Caenorhabditis elegans*. Cold Spring Harbor Press, Cold Spring Harbor, NY) is a small free-living nematode which grows easily and reproduces rapidly in the laboratory. The adult *C. elegans* has about 1000 somatic cells (depending on the sex). The anatomy of *C. elegans* is relatively simple and extremely well-known, and its developmental cell lineage is highly reproducible and completely determined. There are two sexes: hermaphrodites that produce both eggs and sperm and are capable of self fertilization and males that produce sperm and can productively mate with the hermaphrodites.

The self fertilizing mode of reproduction greatly facilitates the isolation and analysis of genetic mutations and *C. elegans* has developed into a most powerful animal model system. In addition, *C. elegans* has a small genome (~10^8 base pairs) whose sequencing is more advanced than that of any other animal.

Genes that encode G-protein subunits in *C. elegans* were identified using probes to sequences conserved in corresponding mammalian genes. So far six Gα genes have been identified including the nematode homologs of mammalian Gαs, Gαo and Gαq/11 as well as three putative Gα proteins that have not yet been assigned to a mammalian subtype class. Gαo, is encoded by the gene *goa-1*. The Gαo protein from *C. elegans* is 80-87% identical to homologous proteins from other species. Mutations that reduce the function of *goa-1* cause behavioral defects in *C. elegans* including hyperactive locomotion, premature egg-laying, inhibition of pharyngeal pumping, male impotence, a reduction in serotonin-induced inhibition of defecation and reduced fertility. Mutations of *goa-1* homologous to the known activating mutations of mammalian Gαs and Gαi2 or overexpression of wild type *goa-1* caused behavioral defects which appear to be opposite to those conferred by reducing *goa-1* function: sluggish locomotion, delayed egg-laying and hyperactive pharyngeal pumping.

egl-10 is a gene from *C. elegans,* originally identified by mutations that cause defects in egg-laying behavior (C. Trent, N. Tsung and H.R. Horvitz (1983) Genetics 104:619-647). The egg-laying defect appears to involve a pair of serotonergic motor neurons (the HSN cells) which innervate vulva muscles in *C. elegans* hermaphrodites (C. Desai, G. Garriga, S.L. McIntire and H.R. Horvitz (1988) Nature 336:638-646; C. Desai and H.R. Horvitz (1989) Genetics 121:703-7212).

Summary of the Invention

We have discovered a new family of proteins involved in the control of heterotrimeric G-protein mediated effects in both mammalian and non-mammalian cells. We disclose sequences which comprise the conserved domains of nine members of this family and methods for identifying additional members. We have named this family of proteins RGS proteins for Regulators of G-protein Signalling.

In general, the invention features substantially pure nucleic acid (for example, genomic DNA, cDNA, RNA or synthetic DNA) encoding an RGS polypeptide as defined below. In related aspects, the invention also features a vector, a cell (e.g., a bacterial, yeast, nematode, or mammalian cell), and a transgenic animal which includes such a substantially pure DNA encoding an RGS polypeptide.

In preferred embodiments, an *rgs* gene is the *egl-10* gene of a nematode of the genus *C. elegans* or the human homolog, *rgs7*. In another preferred embodiment, the RGS encoding nucleic acid cell is in a transformed animal cell. In related aspects, the invention features a transgenic animal containing a transgene which encodes an RGS polypeptide that is expressed in animal cells which undergo G-protein mediated events (for example, responses to neuropeptides, hormones, chemoattractant chemokines, and odor, and synthetic or naturally responses to opiates).

In a second aspect, the invention features a substantially pure DNA which includes a promoter capable of expressing the *rgs* gene in a cell. In preferred embodiments, the promoter is the promoter native to an *rgs* gene. Additionally, transcriptional and translational regulatory regions are preferably native to an *rgs* gene.

In another aspect, the invention features a method of detecting a *rgs* gene in a cell involving: (a) contacting the *rgs* gene or a portion thereof greater than 9 nucleic acids, preferably greater than 18 nucleic acids in length with a preparation of genomic DNA from the cell under hybridization conditions providing detection of DNA sequences having about 30% or greater sequence identity among the amino acid sequences encoded by the conserved DNA sequences of Fig. 3B or the sequences of sequence ID Nos. 2-5 and the nucleic acid of interacting. Preferably, the region of sequence identity used for hybridization is the DNA sequence encoding one of the sequences in the shaded region depicted in Fig. 3B (e.g., the DNA encoding amino acids 1-43 and 92-120 of the EGL-10 fragment shown in Figure 3B (SEQ ID NO: 1)). More preferably, the region of identity is to the DNA encoding the polypeptide sequence delineated by the solid black in Fig. 3B (e.g., amino acids 36-43 and 92-102 of the EGL-10 sequence shown in Fig. 3B). Even more preferably the sequence identity is to the sequences of ID Nos. 1-5. Most preferably, the sequence identity is to the sequences of SEQ ID NOS: 33 or 34. Most preferably, the sequence identity of the nucleic acid sequences being compaired is 50%.

In another aspect, the invention features a method of producing an RGS polypeptide which involves: (a) providing a cell transformed with DNA encoding an RGS polypeptide positioned for expression in the cell (for example, present on a plasmid or inserted in the genome of the cell); (b) culturing the transformed cell under conditions for expressing the DNA; and (c) isolating the RGS polypeptide.

In another aspect, the invention features substantially pure RGS polypeptide. Preferably, the polypeptide includes a greater than 50 amino acid

sequence substantially identical to a greater than 50 amino acid sequence shown in the Fig. 2, open reading frame, more preferably the identity is to one of the conserved regions of homology shown in Fig. 3B (e.g., the sequences 1-43 and 92-120) and, more preferably, 36-43 and 92-102 of SEQ ID NO: 1 and most preferably, the identity is to one of the sequences shown in SEQ ID NOS: 2-5.

In another aspect, the invention features a method of regulating G-protein mediated events wherein the method involves: (a) providing the *rgs* gene under the control of a promoter providing controllable expression of the *rgs* gene in a cell wherein the *rgs* gene is expressed in a construct capable of delivering an RGS protein in an amount effective to alter said G-protein mediated events. The polypeptide may also be provided directly, for example, in cell culture and therapeutic uses. In preferred embodiments, the *rgs* gene is expressed using a tissue-specific or cell type-specific promoter, or by a promoter that is activated by the introduction of an external signal or agent, such as a chemical signal or agent.

In other aspects, the invention features a substantially pure oligonucleotide including one or a combination of the sequences:

5′ GNIGANAARYTIGANTTRTGG 3′, wherein N is G or A; R is T or C; and Y is A, T, or C (SEQ ID NO: 2);

5′ GNIGANAARYTISGITTRTGG 3′, wherein N is G or A; R is T or C; Y is A, T, or C; and S is A or C (SEQ ID NO: 3);

5′ GNTAIGANTRITTRTRCAT 3′, wherein N is G or A; and R is T or C (SEQ ID NO: 4);

5′ GNTANCTNTRITTRTRCAT 3′, wherein N is G or A; and R is T or C (SEQ ID NO: 5);

the *egl-10* DNA shown in Fig. 2A (SEQ ID NO: 27);

ATCAGCTGTGAGGAGTACAAGAAAATCAAATCACCTTCTAAACTAAGTCCCAAGGC CAAGAAGATCTACAATGAGTTCATCTCTGTGCAGGCAACAAAAGAGGTGAACCTGG ATTCTTGCACCAGAGAGGAGACAAGCCGGAACATGTTAGAGCCCACGATAACCTGT TTTGATGAAGCCCGGAAGAAGATTTTCAACCTG (SEQ ID NO: 15);

CAGCTTGTAAATGTGCTCCTGAGCATCTTCGAATGTGTATCGTCCTGGTTCCTTCAC ATTCTGTGTGGTCTTGTCATAACTCTTCGAATCCAAGTTAATGGCACTGGGGGCCCC CGGAGCCAGAAATTCTTGCCATATTTCCTGTACTCGAGAGGGGACCTCTCGGATAG GCCTTTTCTTCAGGTCCTCCACTGCCAA (SEQ ID NO: 16);

CTGGCCTGTGAGGAGTTCAAGAAGACCAGGTCGACTGCAAAGCTAGTCACCAAGG CCCACAGGATCTTTGAGGAGTTTGTGGATGTGCAGGCTCCACGGAGGTGAATATC GATTTCCAGACCCGAGAGGCCACGAGGAAGAACATGCAGGAGCCGTCCCTGACTT GTTTTGATCAAGCCCAGGGAAAAGTCCACAGCCTC (SEQ ID NO: 17);

GAAGCCTGTGAGGATCTGAAGTATGGGGATCAGTCCAAGGTCAAGGAGAAGGCAG AGGAGATCTACAAGCTGTTCCTGGCACCGGGTGCAAGGCGATGGATCAACATAGAC GGCAAAACCATGGACATCACCGTGAAGGGGCTGAGACACCCCCACCGCTATGTGTT GGACGCGGCGCAGACCCACATTTACATGCTC (SEQ ID NO: 18);

CTGGCTTGTGAGGATTTCAAGAAGGTCAAATCGCAGTCCAAGATGGCAGCCAAAGC CAAGAAGATCTTTGCTGAGTTCATCGCGATCCAGGCTTGCAAGGAGGTAAACCTGG ACTCGTACACACGAGAACACACTAAGGAGAACCTGCAGAGCATCACCCGAGGCTG CTTTGACCTGGCACAAAAACGTATCTTCGGGCTC (SEQ ID NO: 19);

GTTGCCTGTGAGAATTACAAGAAGATCAAGTCCCCCATCAAAATGGCAGAGAAGGC AAAGCAAATCTATGAAGAATTCATCCAGACAGAGGCCCCTAAAGAGGTGAACATT GACCACTTCACTAAAGACATCACCATGAAGAACCTGGTGGAACCTTCCCCTCACAG CTTTGACCTGGCCCAGAAAAGGATCTACGCCCTG (SEQ ID NO: 20);

CTGGCCGTCCAAGATCTCAAGAAGCAACCTCTACAGGATGTGGCCAAGAGGGTGG AGGAAATCTGGCAAGAGTTCCTAGCTCCCGGAGCCCCAAGTGCAATCAACCTGGAT TCTCACAGCTATGAGATAACCAGTCAGAATGTCAAAGATGGAGGGAGATACACATT TGAAGATGCCCAGGAGCACATCTACAAGCTG (SEQ ID NO: 21);

CTAGCGTGTGAAGATTTCAAGAAAACGGAGGACAAGAAGCAGATGCAGGAAAAGG CCAAGAAGATCTACATGACCTTCCTGTCCAATAAGGCCTCTTCACAAGTCAATGTG GAGGGGCAGTCTCGGCTCACTGAAAAGATTCTGGAAGAACCACACCCTCTGATGTT CCAAAAGCTCCAGGACCAGATCTTCAATCTC (SEQ ID NO: 22); and

GAGGCGTGTGAGGAGCTGCGCTTTGGCGGACAGGCCCAGGTCCCCACCCTGGTGGA CTCTGTTTACCAGCAGTTCCTGGCCCCTGGAGCTGCCCGCTGGATCAACATTGACA GCAGAACAATGGAGTGGACCCTGGAGGGGCTGCGCCAGCCACACCGCTATGTCCT AGATGCAGCACAACTGCACATCTACATGCTC (SEQ ID NO: 23).

In another aspect, the invention features a substantially pure polypeptide including one or a combination of the amino acid sequences:

Xaa_1 Xaa_2 Xaa_3 Glu Xaa_4 Xaa_5 Xaa_6 Xaa_7, wherein Xaa_1 is I, L, E, or V, preferably L; Xaa_2 is A, S, or E, preferably A; Xaa_3 is C or V, preferably C; Xaa_4 is D, E, N, or K, preferably D; Xaa_5 is L, Y, or F; Xaa_6 is K or R, preferably R; and Xaa_7 is K, R, Y, or F, preferably K (SEQ ID NO: 25); and

Xaa_1 Xaa_2 Xaa_3 Xaa_4 Xaa_5 Xaa_6 Xaa_7 Xaa_8 Xaa_9 Xaa_{10} Lys, wherein Xaa_1 is F or L, preferably F; Xaa_2 is D, E, T, or Q, preferably D; Xaa_3 is E, D, T, Q, A, L, or K; Xaa_4 is A or L, preferably A; Xaa_5 is Q or A, preferably Q; Xaa_6 = L, D, E, K, T, G, or H; Xaa_7 is H, R, K, Q or D; Xaa_8 is I or V, preferably I; Xaa_9 = Q, T, S, N, K, M, G or A (SEQ ID NO: 26). More preferably, the sequences are LACEDXaaK, wherein Xaa is L, Y, or F and (SEQ ID NO: 33) FDXaa,AQ$Xaa_2$$Xaa_3IXaa_4$, wherein Xaa, is E, D, T, Q, A, L, or K; Xaa_2 is L, D, E, K, T, G, or H; and Xaa_3 is H, R, K, Q, or D (SEQ ID NO: 34).

In preferred embodiments the invention features polypeptides having the sequences substantially identical to the EGL-10 and the human RGS2 polypeptides shown in Fig. 3C. More preferably, the polypeptides are identical

to the sequences of EGL-10 and human RGS2 provided in Fig. 3C.

In another aspect, the invention features a method of isolating a *rgs* gene or fragment thereof from a cell, involving: (a) providing a sample of cellular DNA; (b) providing a pair of oligonucleotides having sequence homology to a conserved region of an *rgs* gene (for example, the oligonucleotides of SEQ ID NOS: 2-5); (c) combining the pair of oligonucleotides with the cellular DNA sample under conditions suitable for polymerase chain reaction-mediated DNA amplification; and (d) isolating the amplified *rgs* gene or fragment thereof. Where a fragment is obtained by PCR standard library screening techniques may be used to obtain the complete coding sequence. In preferred embodiments, the amplification is carried out using a reverse-transcription polymerase chain reaction, for example, the RACE method.

In another aspect, the invention features a method of identifying a *rgs* gene in a cell, involving: (a) providing a preparation of cellular DNA (for example, from the human genome); (b) providing a detectably-labelled DNA sequence (for example, prepared by the methods of the invention) having homology to a conserved region of an *rgs* gene; (c) contacting the preparation of cellular DNA with the detectably-labelled DNA sequence under hybridization conditions providing detection of genes having 50% or greater sequence identity; and (d) identifying an *rgs* gene by its association with the detectable label.

In another aspect, the invention features a method of isolating an *rgs* gene from a recombinant DNA library, involving: (a) providing a recombinant DNA library; (b) contacting the recombinant DNA library with a detectably-labelled gene fragment produced according to the PCR method of the invention under hybridization conditions

providing detection of genes having 50% or greater sequence identity; and (c) isolating a member of an *rgs* gene by its association with the detectable label.

In another aspect, the invention features a method of isolating an *rgs* gene from a recombinant DNA library, involving: (a) providing a recombinant DNA library; (b) contacting the recombinant DNA library with a detectably-labelled RGS oligonucleotide of the invention under hybridization conditions providing detection of genes having 50% or greater sequence identity; and (c) isolating an *rgs* gene by its association with the detectable label.

In another aspect, the invention features a recombinant polypeptide capable of altering G-protein mediated events wherein the polypeptide includes a domain having a sequence which has at least 70% identity to at least one of the sequences of sequence ID Nos. 1, 6-14, 25 or 26. More preferably, the region of identity is 80% or greater, most preferably the region of identity is 95% or greater.

In another aspect, the invention features an *rgs* gene isolated according to the method involving: (a) providing a sample of cellular DNA; (b) providing a pair of oligonucleotides having sequence homology to a conserved region of an *rgs* gene; (c) combining the pair of oligonucleotides with the cellular DNA sample under conditions suitable for polymerase chain reaction-mediated DNA amplification; and (d) isolating the amplified *rgs* gene or fragment thereof.

In another aspect, the invention features an *rgs* gene isolated according to the method involving: (a) providing a preparation of cellular DNA; (b) providing a detectably-labelled DNA sequence having homology to a conserved region of an *rgs* gene; (c) contacting the preparation of DNA with the detectably-labelled DNA

sequence under hybridization conditions providing detection of genes having 50% or greater sequence identity; and (d) identifying an *rgs* gene by its association with the detectable label.

In another aspect, the invention features an *rgs* gene isolated according to the method involving: (a) providing a recombinant DNA library; (b) contacting the recombinant DNA library with a detectably-labelled *rgs* gene fragment produced according to the method of the invention under hybridization conditions providing detection of genes having 50% or greater sequence identity; and (c) isolating an *rgs* gene by its association with the detectable label.

In another aspect, the invention features a method of identifying an *rgs* gene involving: (a) providing a mammalian cell sample; (b) introducing by transformation (e.g. biolistic transformation) into the cell sample a candidate *rgs* gene; (c) expressing the candidate *rgs* gene within the cell sample; and (d) determining whether the cell sample exhibits an alteration in G-protein mediated response, whereby a response identifies an *rgs* gene.

Preferably, the cell sample used herein is selected from cardiac myocytes or other smooth muscle cells, neutrophils, mast cells or other myeloid cells, insulin secreting β-cells, COS-7 cells, or xenopus oocytes. In other preferred embodiments the candidate *rgs* gene is obtained from a cDNA expression library, and the RGS response is a membrane trafficking or secretion response or an alteration on [H^3] IP3 or cAMP Levels.

In another aspect, the invention features an *rgs* gene isolated according to the method involving: (a) providing a cell sample; (b) introducing by transformation into the cell sample a candidate *rgs* gene; (c) expressing the candidate *rgs* gene within the tissue sample; and (d) determining whether the tissue sample

exhibits a G-protein mediated response or decrease thereof, whereby a response identifies an *rgs* gene.

In another aspect, the invention features a purified antibody which binds specifically to an RGS family protein. Such an antibody may be used in any standard immunodetection method for the identification of an RGS polypeptide.

In another aspect, the invention features a DNA sequence substantially identical to the DNA sequence shown in Figure 2A. In a related aspect, the invention features a DNA sequence substantially identical to the DNA sequence shown in Fig. 7.

In two additional aspects, the invention features a substantially pure polypeptides having sequences substantially identical to amino acid sequences shown in Figure 3C (SEQ ID NOS:27 and 40).

In another aspect, the invention features a kit for detecting compounds which regulate G-protein signalling. The kit includes RGS encoding DNA positioned for expression in a cell capable of producing a detectable G-protein signalling response. Preferably, the cell is a cardiac myocyte, a mast cell, or a neutrophil.

In a related aspect, the invention features a method for detecting a compound which regulates G-protein signalling. The method includes:

i) providing a cell having RGS encoding DNA positioned for expression; ii) contacting the cell with the compound to be tested; iii) monitoring the cell for an alteration in G-protein signalling response.

Preferably, the cell used in the method is a cardiac myocyte, a mast cell, or a neutrophil, and the responses assayed are an electrophysical response, a degranulation response, or IL-8 mediated response, respectively.

For aforementioned methods involving the use of RGS proteins or *rgs* genes it is noted that the use IR-20/BL34 or gos-8 nucleic acids or proteins encoded there from are also included as methods of the invention. Preferably 1R20/BL34 and gos-8 nucleic and encoded proteins are used in methods for regulating G-proein signalling.

By "*rgs*" is meant a gene encoding a polypeptide capable of altering a G-protein mediated response in a cell or a tissue and which has at least 50% or greater identity to the conserved regions described in Fig. 3B. The preferred regions of identity are as described below under "conserved regions." An *rgs* gene is a gene including a DNA sequence having about 50% or greater sequence identity to the RGS sequences which encode the conserved polypeptide regions shown in Fig. 3B and described below, and which encodes a polypeptide capable of altering a G-protein mediated response. EGL-10 and the human *rgs2* are examples of *rgs* genes encoding the EGL-10 polypeptide from *C.elegans* and a human RGS polypeptide, respectively.

By "polypeptide" is meant any chain of amino acids, regardless of length or post-translational modification (e.g., glycosylation or phosphorylation).

By "substantially identical" is meant a polypeptide or nucleic acid exhibiting at least 50%, preferably 85%, more preferably 90%, and most preferably 95% homology to a reference amino acid or nucleic acid sequence. For polypeptides, the length of comparison sequences will generally be at least 16 amino acids, preferably at least 20 amino acids, more preferably at least 25 amino acids, and most preferably 35 amino acids. For nucleic acids, the length of comparison sequences will generally be at least 50 nucleotides, preferably at

least 60 nucleotides, more preferably at least 75 nucleotides, and most preferably 110 nucleotides.

Sequence identity is typically measured using sequence analysis software (e.g., Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). Such software matches similar sequences by assigning degrees of homology to various substitutions, deletions, substitutions, and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine and tyrosine.

By a "substantially pure polypeptide" is meant an RGS polypeptide which has been separated from components which naturally accompany it. Typically, the polypeptide is substantially pure when it is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight, RGS polypeptide. A substantially pure RGS polypeptide may be obtained, for example, by extraction from a natural source (e.g., a human or rat cell); by expression of a recombinant nucleic acid encoding an RGS polypeptide; or by chemically synthesizing the protein. Purity can be measured by any appropriate method, e.g., those described in column chromatography, polyacrylamide gel electrophoresis, or by HPLC analysis.

A protein is substantially free of naturally associated components when it is separated from those contaminants which accompany it in its natural state.

Thus, a protein which is chemically synthesized or produced in a cellular system different from the cell from which it naturally originates will be substantially free from its naturally associated components. Accordingly, substantially pure polypeptides include those derived from eukaryotic organisms but synthesized in *E. coli* or other prokaryotes.

By "substantially pure DNA" is meant DNA that is free of the genes which, in the naturally-occurring genome of the organism from which the DNA of the invention is derived, flank the gene. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote; or which exists as a separate molecule (e.g., a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. It also includes a recombinant DNA which is part of a hybrid gene encoding additional polypeptide sequence.

By "transformed cell" is meant a cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a DNA molecule encoding (as used herein) an RGS polypeptide.

By "positioned for expression" is meant that the DNA molecule is positioned adjacent to a DNA sequence which directs transcription and translation of the sequence (i.e., facilitates the production of, e.g., an RGS polypeptide, a recombinant protein or a RNA molecule).

By "reporter gene" is meant a gene whose expression may be assayed; such genes include, without limitation, β-glucuronidase (GUS), luciferase, chloramphenicol transacetylase (CAT), and β-galactosidase.

By "promoter" is meant minimal sequence sufficient to direct transcription. Also included in the invention are those promoter elements which are sufficient to render promoter-dependent gene expression controllable for cell-type specific, tissue-specific or inducible by external signals or agents; such elements may be located in the 5' or 3' regions of the native gene.

By "operably linked" is meant that a gene and a regulatory sequence(s) are connected in such a way as to permit gene expression when the appropriate molecules (e.g., transcriptional activator proteins) are bound to the regulatory sequence(s).

By "transgene" is meant any piece of DNA which is inserted by artifice into a cell, and becomes part of the genome of the organism which develops from that cell. Such a transgene may include a gene which is partly or entirely heterologous (i.e., foreign) to the transgenic organism, or may represent a gene homologous to an endogenous gene of the organism.

By "transgenic" is meant any cell which includes a DNA sequence which is inserted by artifice into a cell and becomes part of the genome of the organism which develops from that cell. As used herein, the transgenic organisms are generally transgenic rodents and the DNA (transgene) is inserted by artifice into the genome.

By an "*rgs* gene" is meant any member of the family of genes characterized by their ability to regulate a G-protein mediated response and having at least 20%, preferably 30%, and most preferably 50% amino acid sequence identity to one of the conserved regions of one of the RGS members described herein (i.e., either the *egl-10* gene or the *rgs* 1-9 gene sequences described herein). rgs gene family does not include the FlbA, the Sst-2, CO5B5.7, GOS-8, BL34 (also referred as 1R20) gene sequences.

By "conserved region" is meant any stretch of six or more contiguous amino acids exhibiting at least 30%, preferably 50%, and most preferably 70% amino acid sequence identity between two or more of the RGS family members. Examples of preferred conserved regions are shown (as overlapping or designated sequences) in Figs. 3A and 3B and include the sequences provided by seq ID Nos. 2-5, 25 and 26. Preferably, the conserved region is a region shown by shading blocks in Fig. 3B (e.g., amino acids 1-43 and 92-120 of the EGL-10 sequence shown in Fig. 3B (SEQ ID NO: 1). More preferably, the conserved region is the region delineated by a solid block in Fig. 3B (e.g., amino acids 36-43 and 92-102 of the EGL-10 sequence of Fig. 3B). Even more preferably, the conserved region is defined by the sequences of SEQ ID NOS: 1-5. Most preferably, the sequences are defined by the sequences of SEQ ID NOS: 33 and 34.

By "detectably-labelled" is meant any means for marking and identifying the presence of a molecule, e.g., an oligonucleotide probe or primer, a gene or fragment thereof, or a cDNA molecule. Methods for detectably-labelling a molecule are well known in the art and include, without limitation, radioactive labelling (e.g., with an isotope such as ^{32}P or ^{35}S) and nonradioactive labelling (e.g., chemiluminescent labelling, e.g., fluorescein labelling).

By "transformation" is meant any delivery of DNA into a cell. Methods for delivery of DNA into a cell are well known in the art and include, without limitation, viral transfer, electroportion, lipid mediated transfer and biolistic transfer.

By "biolistic transformation" is meant any method for introducing foreign molecules into a cell using velocity driven microprojectiles such as tungsten or gold particles. Such velocity-driven methods originate from

pressure bursts which include, but are not limited to, helium-driven, air-driven, and gunpowder-driven techniques. Biolistic transformation may be applied to the transformation or transfection of a wide variety of cell types and intact tissues including, without limitation, intracellular organelles, bacteria, yeast, fungi, algae, pollen, animal tissue, plant tissue and cultured cells.

By "purified antibody" is meant antibody which is at least 60%, by weight, free from proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably 90%, and most preferably at least 99%, by weight, antibody, e.g., an EGL-10 specific antibody. A purified RGS antibody may be obtained, for example, by affinity chromatography using recombinantly-produced RGS protein or conserved motif peptides and standard techniques.

By "specifically binds" is meant an antibody which recognizes and binds an RGS protein but which does not substantially recognize and bind other molecules in a sample, e.g., a biological sample, which naturally includes RGS protein.

By "regulating" is meant conferring a change (increase or decrease) in the level of a G-protein mediated response relative to that observed in the absence of the RGS polypeptide, DNA encoding the RGS polypeptide, or test compound. Preferably, the change in response is at least 5%, more preferably, the change in response is greater than 20%, and most preferably, the change in response level is a change of more than 50% relative to the levels observed in the absence of the RGS compound or test compound.

By "G-protein signalling response" is meant a response mediated by heterotrimeric guanine nucleotide

binding proteins. It will be appreciated that these responses and assays for detecting these responses are well-known in the art. For example, many such responses are described in the references provided in the detailed description, below.

By an "effective amount" is meant an amount sufficient to regulate a G-protein mediated response. It will be appreciated that there are many ways known in the art to determine the effective amount for a given application. For example, the pharmacological methods for dosage determination may be used in the therapeutic context.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

Detailed Description

The drawings will first be described.

Drawings

Fig. 1A is the genetic map of region of *C. elegans* chromosome V that contains the gene *egl-10*.

Fig. 1B is a physical map of the *egl-10* region of the *C. elegans* genome.

Fig. 2A is the nucleotide sequence of *egl-10* cDNA and the amino acid sequence from the open reading frame, EGL-10 (SEQ ID NO: 27. ADD SEQ NO for egl-10 cDNA).

Fig. 2B shows the positions of *egl-10* introns and exons and the positions of *egl-10* mutations therein.

Fig. 2C is Northern Blot analysis with *egl-10* cDNA.

Fig. 2D is the sequence of *egl-10* mutations.

Fig. 3A is a diagram of EGL-10 and structurally related proteins showing amino acid sequences in conserved domains.

Fig. 3B shows the sequences of RGS regions of homology (SEQ ID NOS: 1, 6-14, 28-32, 30-32, and 36-39. The RGS-3-4 sequences are isolated from the rest).

Fig. 3C is a comparison of the EGL-10 amino acid sequence and the human RGS7 sequence (SEQ ID NOS 27 and 40).

Fig. 4 is a photograph of a Northern blot showing distribution of *egl-10* homolog mRNAs in various rat tissues. Fig. 5 shows the partial DNA sequences from the rat *rgs* genes, referred to as RGS5 1-7 sequences (SEQ ID NOS: 15-23).

Fig. 6A - 6G show EGL-10 protein expression. Fig. 6A shows western blot analysis of protein extracts from wild-type and *egl-10(md176)* worms probed with the affinity purified anti-EGL-10 polyclonal antibodies. The filled arrow indicates the position of the EGL-10 protein detected in wild-type but not in *egl-10* mutant extracts. The open arrow indicates the 47 kD protein that cross-reacted with the EGL-10 antibodies but was not a product of the EGL-10 gene. The positions of molecular weight markers are indicated, with their sizes in kD. Fig. 6B shows anti-EGL-10 antibody staining of the head of a wild-type adult hermaphrodite. The dark immunoperoxidase stain labeled the neural processes of the nerve ring (arrow). Fig. 6C shows anti-EGL-10 antibody staining of the head of an *egl-10(md176)* adult hermaphrodite, prepared in parallel to the preparation on Fig. 6B and lacking any specific staining. Fig. 6D shows anti-EGL-10 immunofluorescence staining in the mid-body region of a wild-type adult. The fluorescence here and in panels E-G appears white on a black background, the reverse of the staining in Fig. 6B and 6C. The arrow points to the brightly stained ventral cord neural processes. Body-wall muscle cells on either side of the ventral cord contained brightly

stained spots arranged in linear arrays. Body-wall muscles throughout the animal showed similar staining. Fig. 6E shows fluorescence in the head of a transgenic adult carrying a fusion of the *egl-10* promoter and N-terminal coding sequences to the green fluorescent protein (GFP) gene. The fusion protein is localized in spots within the body-wall muscles similar to those seen in Fig. 6D. GFP fluorescence was also present in neural processes and cell bodies out of the plane of focus. Fig. 6F shows anti-EGL-10 antibody staining in the head of a transgenic worm carrying the *nIs51* multicopy array of wild-type *egl-10* genes. Fig. 6G shows anti-EGL-10 antibody staining in the vulva region of *nIs51* worms. The open arrow points to the vulva. The large filled arrow indicates the HSN neuron. The small filled arrow points to the ventral cord and associated neural cell bodies.

Fig. 7 shows the human *rgs2* cDNA sequence (SEQ ID NO:41)

I. <u>EGL-10 identifies a new family of heterotrimeric G-protein pathway associated proteins which are regulators of G-protein signalling (RGS's)</u>.

A. *Characteristics of egl-10.*

1. *Phenotypes conferred by mutation of the egl-10 gene.*

The phenotypes conferred by mutations in *egl-10* have been further characterized. As previously described, *egl-10* loss-of-function mutants fail to lay eggs and have sluggish locomotory behavior (C. Trent, et al. (1983) Genetics <u>104</u>:619-647)). We have now discovered that the overexpression of *egl-10* produces the opposite effects: hyperactive egg-laying and locomotion. More generally, we have discovered that the rates of egg-laying and

locomotory behaviors are proportional to the number of functional copies of *egl-10*.

The phenotypes conferred by mutations in *egl-10* are strikingly similar to those conferred by mutations in *goa-1* (J.E. Mendel, et al. (1995) Science 267:1652-5); L. Ségalat, et al. (1995) Science 267:1648-52). However, these phenotypes are reversed relative to the level of gene function: mutations of *egl-10* which enhance gene function increase the rate of various behaviors whereas those mutations that reduce gene function decrease the rates of these behaviors. By contrast, mutations *goa-1* which reduce function increase the rate of behaviors, whereas overexpression decreases the rate of the behaviors. The occurrence of such a similar constellation of phenotypes strongly suggests that the functions of EGL-10 and GOA-1 proteins have related functions, components of the same or parallel genetic pathway. Since GOA-1 is the nematode homolog of the heterotrimeric G-protein, Gαo, it is thus likely that EGL-10 plays a role in one or more heterotrimeric G-protein regulatory pathways which contains Gαo.

We have further discovered that loss of function mutations in *egl-10* confer resistance to drugs that effect *C. elegans* by acting as inhibitors of acetylcholinesterase (AChE). Other mutations that confer resistance to AChE inhibitors have been shown to reduce the synthesis and packaging of the neurotransmitter acetylcholine (ACh) or to reduce the function of genes that encode proteins that comprise the biochemical machinery responsible for neurotransmitter release (M. Nguyen, A. Alfonso, C.D. Johnson and J.B. Rand (1995) Genetics 140:527-35). This result indicates that EGL-10, and presumably its associated G-protein coupled pathways, function to modulate the release of acetylcholine in *C.*

elegans and may be involved in the release of other neurotransmitters as well.

2. The cloning and sequencing of the *egl-10* gene.

egl-10 had been previously mapped between *rol-4* and *lin-25* on chromosome V. Additional mapping, using RFLP markers, placed *egl-10* within ~15Kb of DNA, contained entirely on a single cosmid clone (Fig. 1A). Germline transformation with DNA from a subclone from the region rescues the phenotype conferred by a mutation that reduces *egl-10* function. Furthermore, the rescue is blocked by insertion of a synthetic oligonucleotide which interrupts an open reading frame, located entirely within the rescuing fragment, with a stop codon (Fig. 1B). The open reading thus very likely encodes the EGL-10 protein.

The fragment used for transformation rescue was used to screen several *C. elegans* cDNA libraries. The longest cDNA obtained (3.2 kb) was sequenced on both strands. The cDNA was judged to be full length since it contains a sequence matching the *C. elegans* trans-spliced-leader SL1 (M. Krause and D. Hirsh (1987) Cell 49:753-61). The regions of the genomic clone to which this cDNA hybridized were sequenced on one strand. The *egl-10* genomic structure was deduced by comparing the cDNA and genomic sequences. The 3169 nucleotide long sequence obtained from the cDNA and the 555 amino acid long predicted amino acid sequence of the putative EGL-10 protein are shown in Fig. 2A. The organization of exons and introns within genomic DNA are shown in Fig. 2B. Northern blot analysis (Fig. 2C) showed the presence of a single mRNA species at ~3.2kB.

We sequenced the putative *egl-10* genomic cDNA obtained from a collection of independently isolated *egl-10* mutations. Nine mutations induced by chemical mutagenesis were shown to contain point mutations within

the gene. Six of the mutations created new stop codons leading to truncated proteins; the other three mutations produced amino acid sequence changes (Fig. 2D). Five spontaneous *egl-10* mutations, isolated from a genetically unstable strain of *C. elegans*, were shown to contain either an insertion of the transposon Tc1 or a rearrangement (Fig. 2D). Locations of these mutations within the gene are shown in Figures 2A and 2B. The observation that many *egl-10* mutations have detectable defects in a putative *egl-10* cDNA is considered proof that this cDNA encodes the EGL-10 gene product.

B. egl-10 is a member of a new gene family - rgs family.

The *egl-10* gene consists largely of novel sequences. However, a search of protein sequence databases indicated that the gene encodes a 119 amino acid domain (Figure 3A) that is also present in the predicted amino acid sequences of two small human genes, known as BL34/IR20 and GOS-8. The functions of BL34/1R20 and GOS-8 were previously completely unknown, and these genes were identified only as sequences whose expression is increased in B lymphocytes stimulated with phorbol esters. In addition, a conceptual gene of unknown function, called *C05B5.7*, identified by the *C. elegans* genome sequencing project, also contains this conserved domain. Thus, EGL-10 appears to identify a family of proteins with multiple members in the same species and homologs in related species. By using degenerate probes from the conserved domain (in EGL-10, BL34/1R20, GOS-8, and *C05B5.7*) and PCR, we isolated 9 novel sequences that contain the conserved domain from rat brain cDNA (labelled as rat gene fragments 3 through 11; Fig. 3B). The rat gene fragments isolated using this method are called *rgss-1* through *rgss-9* for regulator G-protein

signalling similarity. It appears that there exists a substantial number of genes in mammals that are members of the *rgs* family.

We also observed weak sequence similarities between portions of the conserved domain in *egl-10* and regions of the *sst-2* gene of the yeast *Saccharomyces cerevisiae* and the *flbA* gene is the fungus *Aspergillis nidulans*. The function of the SST-2 protein appears to involve one mode of adaptation in the G-protein pathway responsible for transduction of the binding of the yeast mating factors a and α to their respective 7-TMRs. Evidence from studies of the sensitivity of yeast Gα to a specialized form of proteolysis, suggests that SST-2 protein may interact directly with Gα. The functions of FlbA are much less well studied.

II. Methods for identifying new members of the *rgs/egl-10* gene family.

The region of homology we have identified may be used to obtain additional members of the RGS family. For example, sequences from the genes *rgss-1* through *rgss-9* were obtained by PCR using degenerate oligonucleotide primers designed to encode the amino acid sequences of EGL-10, 1R20, and BL34 proteins at the positions indicated in Fig. 3B. Two 5' primers pools were used with two 3' primer pools in all four possible combinations. After two rounds of amplification all four primer pairs gave a detectable products of ~240 bp. These products were used to prepare clone libraries, restriction maps were prepared for selected clones from each library, clones with different restriction maps were divided into classes, and then several clones from each restriction map class were sequenced. In total 47 clones were sequenced. Each of the nine *rgs* genes identified by this approach was isolated at least twice. As a result,

we conclude that it is likely that we have identified nearly all the *rgs* genes that can be amplified from rat brain cDNA using these primer pairs.

At least some of the *rgs* sequences are expressed in a wide variety of mammalian tissues, as demonstrated by Northern blotting (Fig. 4). Additional G-protein signalling genes may be identified by using the same primer pairs with cDNA from other rat tissues, with human cDNAs or with cDNAs from other species. In addition, additional *rgs* genes may be identified using alternate primers, based on different amino acid sequences that are conserved not only in the EGL-10, BL34 and 1R20 proteins, but also in the conceptual protein encoded by C05B5.7, in SST2 and FlbA and in the proteins encoded by the *rgs* genes described herein.

III. The functional characterization of new *rgs*/RGS family members

A. General considerations.

The function of newly discovered *rgs* genes can be determined by analyzing:
i) the effects of RGS proteins *in vivo* and *in vitro*,
ii) the effects of antibodies specific to RGS proteins,
or iii) the effects of antisense *rgs* oligonucleotides
in well characterized assay systems that measure functions of mammalian heterotrimeric G-protein coupled pathways. Relevant assays for RGS activity include systems based on responses of intact cells or cell lines to ligands that bind to 7-TMRs, systems based on responses of premeabilized cells and cell fragments to direct or indirect activation of G-proteins and *in vitro* systems that measure biochemical parameters indicative of the functioning of G-protein pathway components or an interaction between G-protein pathway components. The G-protein pathway components whose functions or

interactions are to be measured can be produced either through the normal expression of endogenous genes, through induced expression of endogenous genes, through expression of genes introduced, for example, by transfection with a virus that carries the gene or a cDNA for the gene of interest or by microinjection of cDNAs, or by the direct addition of proteins (either recombinant or purified from a relevant tissue) to an *in vitro* assay system.

B. Specific assay systems which may be employed to detect and screen new RGS genes and polypeptides.

Specific assay systems, including those which are relevant to the pathophysiology of human disease and/or are useful for the discovery and characterization of new targets for human therapeutics are as follows:

1. Assays based on natural responses of intact cells.

Many mammalian cells, for example cardiac myocytes, other smooth muscle cells, neutrophils, mast cells and other classes of myeloid cells and insulin secreting β cells of the pancreas have readily detected responses mediated by heterotrimeric G-protein dependent pathways. To determine if a particular RGS protein is involved in such a pathway, one may compare the response of normal cells to the response which is obtained in cells transfected or transiently transformed by the *rgs* gene. Transformation may be done with the RGS cDNA under the appropriate promotor or with a construct designed to overexpress antisense oligonucleotides to the *rgs* mRNA.

For example, we could express an *rgs* gene or antisense oligonucleotides to an *rgs* mRNA in mammalian cardiac myocytes as described, for example, by Ramirez et al. (M.T. Ramirez, G.R. Post, P.V. Sulakhe and J.H. Brown (1995) J. Biol. Chem. 270:8446-51). Cardiac

myocytes system respond to a variety of ligands, for example α- and β-adrenergic agonists and muscarinic agonists, by altering membrane conductances, including conductances to Cl^-, K^+ and Ca^{2+}. These effects are mediated by G-proteins through a web of both second messenger mediated and membrane delimited effects and are readily measured with a variety of well known electrophysiological technologies (for example: T.C. Hwang, M. Horie, A.C. Nairn and D.C. Gadsby (1992) J. Gen. Physiol. 99:465-89.). We would compare the response of normal myocytes to cells that overexpress a particular *rgs* gene or antisense oligonucleotides to a particular *rgs* mRNA. If no difference was observed, we would conclude that the particular RGS protein played no detectable role in cardiac myocyte physiology. On the other hand, if alterations in membrane currents were observed we would dissect the altered response using pharmacology, permeabilized cell systems and reconstitute G-protein pathways systems to determine the site of action of the RGS protein. One may use this system for specific screens to identify and test compounds that mimic or block the function of the RGS protein.

2. Assays based on expression of cloned genes in particular cells or cell lines.

The involvement of a RGS protein in some known functions and interactions between components of heterotrimeric G-protein pathways can be efficiently assessed in model systems designed for easy and efficient overexpression of cloned genes. One well developed system uses COS-7 cells (monkey kidney cells which possess the ability to replicate SV-40 origin-containing plasmids) as a host for the expression of cloned genes and cDNAs (D.Q. Wu, C.H. Lee, S.G. Rhee and M.I. Simon (1992) J. Biol. Chem. 267:1811-7). Recently, for example,

overexpression of G-protein pathway genes in COS-7 cells was used to determine the capability of two forms of interleukin-8 receptor to activate the 5 different Gα subunits of the Gq family by measuring subsequent effects on the activity of two alternate types of PI-PLCβ, measured by quantified the formation of $[H^3]$IP3 in cells prelabelled with radioactive inositol (D. Wu, G.J. LaRosa and M.I. Simon (1993) Science 262:101-3). Similarly co-expression in COS-7 cells has been used to quantitate the effects of proteins that inhibit signalling by activated G-proteins (W.J. Koch, B.E. Hawes, J. Inglese, L.M. Luttrell and R.J. Lefkowitz (1994) J. Biol. Chem. 269:6193-7).

A useful alternative to cells lines, more amenable to the study of membrane delimited activation of ion channels involves the transient production of proteins following injection of mRNAs into *Xenopus* oocytes (E. Reuveny, P.A. Slesinger, J. Inglese, J.M. Morales, J.A. Iniguez-Lluhi, R.J. Lefkowitz, H.A. Bourne, Y.N. Jan and L.Y. Jan (1994) Nature 370:143-6). For example, the coexpression of two 7-TMRs (serotonin type 1C receptor and thyrotropin releasing hormone receptor) may be coupled with overexpression of one of seven alternate Gα subunits and with one of two alternate PI-PLCβs or adenylyl cyclase and the cystic fibrosis transmembrane conductance regulator (CFTR) (M.W. Quick, M.I. Simon, N. Davidson, H.A. Lester and A.M. Aragay (1994) J. Biol. Chem. 269:30164-72). Combined with expression of antisense oligonucleotides designed to block endogenous pathways, these systems can be engineered to measure specific interactions between 7-TMRS, G subunits, effectors, various inhibitors as well as components controlled by effectors. To determine the effect of an RGS protein one may compare the effect in transfected COS-7 cells or Xenopus oocytes with and without

cotransfection with the *rgs* gene or cDNA, one may also transfect an *rgs* gene construct designed to overexpress antisense oligonucleotides to endogenous *rgs* mRNAs.

If a RGS protein-dependent alteration of a G-protein dependent response is observed, one may utilize pharmacological tools and reconstitute G-protein pathways systems to determine the site of action of the RGS protein. From these experiments, a specific screen for identifying and testing compounds that mimic or block the function of the RGS protein may be developed.

3. Assays utilizing premeabilized cells.

The role of RGS proteins in intracellular events such as membrane trafficking or secretion can be studied in systems utilizing permeabilized cells, such as mast cells (T.H. Lillie and B.D. Gomperts (1993) Biochem. J. 290:389-94), chromaffin cells of the adrenal medulla (N. Vitale, D. Aunis and M.F. Bader (1994) Cell. Mol. Biol. 40:707-15) or more highly purified systems derived from these cells (J.S. Walent, B.W. Porter and T.F.J. Martin (1992) Cell 70:765-775). The determine the effects of RGS proteins one may compare the extent and kinetics of GTP or γS-GTP induced secretion in the presence and absence of excess RGS protein or antibodies specific to RGS proteins.

If an RGS protein-dependent alteration of membrane trafficking or secretion is observed, further experiments may be used to explore the specificity and generality of this action and to determine the precise site of action of the RGS protein. From these experiments, a specific screen for identifying and testing compounds that mimic or block the function of the RGS protein can be constructed.

4. Assays utilizing reconstituted G-protein pathways.

The ability to assess specific protein-protein interactions between specific components that function within G-protein pathways may be employed to assign RGS functions. These assays generally use recombinant proteins purified from an efficient expression systems, most commonly, i) insect Sf9 cells infected with recombinant baculovirus or ii) *E. coli*. Specific interactions which form part of G-protein pathways are then reconstituted with purified or partially purified proteins. The effects of RGS proteins on such systems can be easily assessed by comparing assays in the presence and absence of excess RGS protein or antibodies specific to RGS proteins. From these experiments, specific screens for identifying and testing compounds that mimic or block the function of the RGS protein can be developed.

Uses

RGS DNA, polypeptides, and antibodies have many uses. The following are examples and are not meant to be limiting. The RGS encoding DNA and RGS polypeptides may be used to regulate G-protein signalling and to screen for compounds which regulate G-protein signalling. For example, RGS polypeptides which increase secretion may be used industrially to increase the secretion into the media of commercially useful polypeptides. Once proteins are secreted, they may be more readily harvested. One method of increasing such secretion involves the construction of a transformed host cell which synthesizes both the RGS polypeptide and the commercially important protein to be secreted (e.g, TPA). RGS proteins, DNA, and antibodies may also be used in the diagnosis and treatment of disease. For example, regulation of G-

protein signalling may be used to improve the outcome of patients with a wide variety of G-protein related diseases and disorders including, but not limited to: diabetes, hyperplasia, psychiatric disorders, cardiovascular disease, McCune-Albright Syndrome, and Albright hereditary osteopathy.

IV. Deposit Information.

Genebank accession numbers for the sequences provided herein are as follows: The worm sequence, *egl-10;* has number U32326. The rgs sequence fragments isolated from the rat as follows: *rgs5*, U32434; *rgs1*, U32327; *rgs6*, U32435; *rgs7*, U32436; rat *rgs2*, U32328; *rgs3*, U32432; *rgs4*, U32433; *rgs8*, U32437; *rgs8*, U32438. Accession numbers for representative expressed sequence tags from human *rgs* genes are: RGS-1, R12757, F07186; RGS6, D31257, R35272; RGS10, R35472, T57943; RGS13, T94013; RGS11, R11933; RGS12, T92100. The human RS7 accession number is 442439.

V. Examples.

A. Characteristics of egl-10.

1. Nematode strains.

Nematode strains were maintained and grown at 20°C as described by Brenner (Brenner, (1974) Genetics 77:71-94). Genetic nomenclature follows standard conventions (Horvitz et al., (1979) Mol. Gen. Genet. 175:129-33). The following mutations were used: *goa-1(n363, n1134)* (Ségalat et al., (1995) Science 267:1648-51), *arDf1* (Tuck and Greenwald, (1995) Genes & Development 9:341-57), *egl-10* alleles (Trent et al., (1983) Genetics 104:619-47); Desai and Horvitz, (1989) Genetics 121:703-21), *nIs51* (this work), *nIs67* (this work). We also used the following marker mutations, described by Wood (Wood, ed. (1988) Cold Spring Harbor,

New York: Cold Spring Harbor Laboratory): (LG I), *unc-13(e1091)*; (LGV), *unc-42(e270)*, *lin-25(n545)*, *him-5(e1467)*; (LGX), *lin-15(n765)*.

2. The genetic map position of egl-10.

egl-10 had previously been mapped between *rol-4* and *lin-25* on chromosome V (Trent et al., (1983) Genetics 104:619-647; Desai and Horvitz, (1989) Genetics 121:703-21). We characterized four Tc1 transposon insertions found in this interval in the Bergerac strain of *C. elegans*, but not in the standard Bristol (N2) strain: *nP63*, *nP64*, *arP4* and *arP5* (first identified by Tuck and Greenwald, (1995) Genes & Development 9:341-57). From heterozygotes of the genotype *egl-10(n692)/rol-4(sc8) nP63 nP64 arP4 arP5 lin-25(n545) him-5(e1467)*, Rol non-Lin recombinants were selected. Strains homozygous for the recombinant chromosomes were assayed for the Egl-10 phenotypes (sluggish movement and defective egg-laying), and for the presence of each of the transposons by probing Southern blots of genomic DNA with appropriate genomic clones. Nine recombination breakpoints were thus found to distribute as follows: *rol-4* (2/9) *nP63* (0/9) *nP64* (1/9) *egl-10* (1/9) *arP4* (1/9) *arP5* (4/9) *lin-25*. These data place the egl-10 gene in the interval between *nP64* and *arP4* (Figure 1A).

3. goa-1; egl-10 double mutants.

goa-1; egl-10 strains were constructed by using the *unc-13(e1091)* mutation, which lies within 80 kb of the *goa-1* gene (Maruyama and Brenner, (1991) Proc. Nat'l. Acad. Sci. USA 88:5729-33), to balance the *goa-1* mutations. *unc-13/+; egl-10/+* males were mated to *goa-1* hermaphrodites and hermaphrodite cross progeny were placed individually on separate plates. *unc-13/goa-1; egl-10/+* animals were recognized as segregating 1/4 Unc (uncoordinated) and ~1/4 Egl (egg-laying defective) progeny. Among these progeny, Egl non-Unc animals were picked to separate plates, and were judged to be of genotype *goa-1/unc-13; gl-10* if they segr gated 1/4 Unc

and >3/4 Egl progeny. Non-Unc progeny were picked individually to separate plates, and *goa-1; egl-10* animals were recognized as never segregating Unc progeny. The following double mutant strains were constructed: MT8589 *goa-1(n1134); egl-10(n990)*, MT8593 *goa-1(n363); egl-10(n990)*, MT8641 *goa-1(n363); egl-10(n944)*, MT8587 *goa-1(n1134); egl-10(n944)*, *goa-1(n363); egl-10(md176)*.

Animals with reduction of function mutations in both *goa-1* and *egl-10* display a behavioral phenotype that is very similar to that of strains with mutations in *goa-1* alone, i.e. the animals have hyperactive locomotion and precocious egg-laying. This observation implies that EGL-10 protein acts either before or at the same step in the G-protein regulatory pathway as the GOA protein, Gαo.

4. Germline transformation and chromosomal integration of egl-10 transgenes.

Germline transformation (Mello et al., (1991) Embo. J. 10:3959-70) was performed by coinjecting the experimental DNA (80 μg/ml) and the *lin-15* rescuing plasmid *pL15EK* (Clark et al., (1994) Genetics 137, 987-97) into animals carrying the *lin-15(n765)* marker mutation. Transgenic animals typically carry coinjected DNAs as semistable extrachromosomal arrays (Mello et al., (1991) Embo. J. 10:3959-70) and are identified by rescue of the temperature sensitive multivulva phenotype conferred by the *lin-15(n765)* mutation. For *egl-10* rescue experiments, animals of the genotype *egl-10(n692); lin-15(n765)* were injected, and transgenic lines were considered rescued if >90% of the non-multivulva animals did not show the egg laying defective phenotype conferred by the *egl-10(n692)* mutation. Plasmid *pMK120* contains a 15 kb SmaI-FspI fragment of cosmid W08H11, containing the entire *egl-10* gene, into which the self-annealed oligonucleotide 5'-GTGCTAGCACTGCA-3' (SEQ ID NO: 35) was inserted at the unique PstI site, thus disrupting the open reading frame of the fourth *egl-10* exon. *pMK121* was generated by digesting *pMK120* with PstI and ligating,

thus precisely removing the oligonucleotide and restoring the *egl-10* open reading frame. *egl-10* was rescued in all 13 transgenic lines carrying *pMK121* that were generated, while 0/17 *pMK120* lines showed *egl-10* rescue of even a single animal (Fig. 1B).

5. egl-10 cDNAs and the egl-10 genomic structure.

An 8.5 kb ApaI-MscI fragment, encompassing the middle half of the *egl-10* rescuing genomic clone *pMK120*, was used to screen $3.7X10^6$ plaques from four different *C. elegans* cDNA libraries (Barstead and Waterston, (1989) J. Biol. Chem. 264:10177-85; Maruyama and Brenner, (1992) Gene 120:135-41.; Okkema and Fire, (1994) Development 120:2175-86.). Thirteen *egl-10* cDNAs were isolated, the longest of which was 3.2 kb. This cDNA was completely sequenced on both strands using an ABI 373A DNA sequencer (Applied Biosystems, Inc.). The sequence data was compiled on a Sun workstation running software as described by Dear and Staden (Dear and Staden, (1991) Nucleic Acids Research 19:3907-11) and displayed in Fig. 2A. The regions of the *pMK120* genomic clone to which this cDNA hybridized were also sequenced on one strand, and the *egl-10* genomic structure was deduced by comparing the cDNA and genomic sequences (Fig. 2B). The 3.2 kb cDNA was judged to be full length since it contains a sequence matching the *C. elegans* trans-spliced leader SL1 (Krause and Hirsh, (1987) Cell 49: 753-61) at its 5' end, a poly(A) tract at its 3' end (although it lacks a consensus poly(A) addition signal), and matches the length of the 3.2 kb RNA detected by Northern hybridization (Figure 2C). Other cDNAs were shorter but colinear with the 3.2 kb cDNA clone as judged by restriction mapping and end sequencing.

6. egl-10 mutant DNAs.

egl-10 genomic DNA was PCR amplified from *egl-10* mutants in ~1 kb sections using primers designed from the *egl-10* genomic sequence. The PCR products were electrophores d on agarose gels, and the excised PCR

fragments were purified from the agarose by treatment with β-agarase (New England Biolabs) and isopropanol precipitation. The purified PCR products were directly sequenced using the primers that were used to amplify them, as well as primers that annealed to internal sites. Any differences from the wild-type sequence were confirmed by reamplification and resequencing of the site in question. In this way the entire *egl-10* coding sequence as well as sequence 20 bp into each *egl-10* intron was determined for each of ten ethyl methanesulphonate (EMS)-induced *egl-10* alleles (Trent et al., (1983) Genetics 104:619-647; Desai and Horvitz, (1989) Genetics 121:703-21), as well as for the spontaneous allele *md1006*. The alterations discovered are listed in Fig. 2D. One EMS-induced *egl-10* allele, *n953*, appeared to contain no alterations from wild type in the region sequenced, but may contain alterations in other parts of the gene. *md1006* contains no sequence alterations from wild type other than the insertion of a Tc1 transposon at codon 515.

Genomic DNA from each of five spontaneous *egl-10* alleles was analyzed by Southern blotting and probing with clones spanning the *egl-10* gene. *md1006* contains a 1.6 kb insert relative to wild type which was shown to be a Tc1 transposon insertion by PCR amplification using primers that anneal to the Tc1 ends with primers that anneal to *egl-10* sequences flanking the insertion site, and by further sequencing these PCR products. The four other spontaneous alleles each contain multiple restriction map abnormalities spanning the entire *egl-10* locus, and each failed to give PCR amplification products using one or more primer pairs from the *egl-10* gene. None of these alleles appear to be due to a simple insertion or deletion, and we suspect more complex rearrangements may have occurred.

7. Localization of EGL-10 protein in neural processes and subcellular regions of body wall muscle cells.

We raised polyclonal antibodies against recombinant EGL-10 protein. When affinity-purified, these antibodies recognized two major proteins on western blots of total C. elegans proteins (Fig. 6A). The larger of these proteins is the product of the *egl-10* gene, since this protein was absent from extracts of the *egl-10* null mutant *md1176* (Fig. 6A), as well as from extracts of 12 other *egl-10* mutants. This larger protein was detected at a reduced abundance in the weak *egl-10* mutant *n480* and was present at normal abundance in *egl-10(n1125)* animals, which carry a missense mutation that alters amino acid 446. The 47 kD protein recognized by the anti-EGL-10 antibodies is not affected by *egl-10* mutations and thus is not encoded by the *egl-10* gene (Fig. 6A).

We stained wild-type and *egl-10* mutant worms with the affinity-purified anti-EGL-10 antibodies. We observed staining in the nerve ring (Fig. 6B), ventral nerve cord (Fig. 6D), and dorsal nerve cord (not shown) of wild-type animals, but saw no neural staining in *egl-10* mutants (Fig. 6C). The stained structures consisted of bundles of neural processed and were at the locations of the majority of the chemical synapses in the animal (White et al., Phil. Trans. R. Soc. Lond. B 314:1-340, 1986). In neurons EGL-10 protein appeared to be localized exclusively to processes; no staining was seen in the neural cell bodies of wild-type animals. Animals at all stages of development from first-stage larvae to adults showed similar staining of neural processes. The localization of EGL-10 protein to structures in which chemical synapses are made is consistent with a role for EGL-10 in intercellular signalling.

We also used the EGL-10 antibodies to stain worms that overexpress EGL-10 from a multicopy array of *egl-10* transgenes (Figs. 6F, 6G). EGL-10 was detected in neural cell bodies as well as neural processes of these animals, either because overexpression raised the level of EGL-10 protein in cell bodies above the threshold of detection or because ov rexpression of EGL-10 exceeded the capacity

of neurons to localize the protein to processes. Figure. 6F shows that a large number of neurons in the major ganglia of the head region expressed EGL-10. In addition, our examination of the ventral cord neurons, lateral neurons, and tail ganglia suggested that most if not all neurons in C. elegans expressed EGL-10. In particular, the HSN motor neurons, which control egg-laying behavior and appear to be functionally defective in *egl-10* mutants, expressed EGL-10 (Fig. 6F).

A second staining pattern present in wild-type animals consisted of spots arranged in linear arrays within the body-wall muscle cells (Fig. 6D). Although this staining was not absent from *egl-10* null mutants, we nevertheless believe that the EGL-10 protein is localized to these muscle structures, since the muscle stain was more intense in EGL-10 overexpressing animals and was reproduced by *egl-10::gfp* transgenes (see below). The residual antibody stain seen in the muscles of *egl-10* mutants may have been caused by the presence of a cross-reactive protein (perhaps the 45 kD protein detected in our western blots) that is colocalized with EGL-10. The body-wall muscles are used in locomotion behavior (Wood et al., The Nematode *Caenorhabditis elegans*, Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press, 1988), the frequency of which is controlled by *egl-10*. Every body wall muscle cell stained, but no staining was detected in other types of muscle cells, even in animals overexpressing EGL-10. The body-wall muscle stain superimposed on structures visible in Nomarski optics called dense bodies, which function as attachment sites between the body-wall muscles and the cuticle that surrounds them (Wood et al., supra). Each dense body is flanked by membranes of the sarcoplasmic reticulum, and our observations at the light microscope level cannot distinguish between localization of the stain to the dense bodies or to the sarcoplasmic reticulum. The significance of the localization of EGL-10 to these structures is unclear.

Transgenic animals carrying fusions of the *egl-10* promoter and N-terminal coding sequences to the fluorescent reporter protein GFP (Chalfie et al., Science 263:802-805, 1994) showed GFP fluorescence in body-wall muscle cells in the same pattern seen in animals stained with the EGL-10 antibody (Fig. 6E). These experiments demonstrated that the N-terminal 122 amino acids of EGL-10, when fused to GFP, were sufficient to localize the fusion protein to the dense body-sarcoplasmic reticulum-like structures. The EGL-10::GFP fusion proteins were also expressed in neurons but, like overexpressed full-length EGL-10 protein, were not tightly localized to processes, preventing us from identifying the regions of EGL-10 responsible for localization of EGL-10 to neural process.

8. EGL-10 is similar to Sst2p, a negative regulator of G protein signalling in yeast.

The 555 amino acid EGL-10 protein contains a 120-amino acid region near its carboxy-terminus with similarity to several proteins in the sequence databases (Fig. 3A). The similarities with the C. elegans C05B5.7 protein and the BL34/1R20 and G0S8 proteins extend across the entire 120-amino acid region; this region is 34-55% identical in pairwise comparisons among EGL-10 and these other proteins. An additional C. elegans protein, C29H12.3, consists almost entirely of two highly diverged repeats of this domain. The first 43 and last 29 amino acids of the conserved 120-amino acid region are similar to sequences found in the yeast protein Sst2P and the Aspergillus nidulans protein FlbA. Sst2p and FlbA are 30% identical to each other over their entire lengths and show higher conservation in several short regions (Fig. 3A); it is two of these more highly conserved regions that show similarity to the conserved domain found in EGL-10, C05B5.7, BL34/IR20, GOS8 and C29H12.3. Alignments of all of these conserved sequences are shown in Fig. 3B. This figure also shows alignments with the

sequences of nine additional mammalian EGL-10 protein homologs whose isolation is described below.

The similarity of EGL-10 to Sst2p is of particular interest, since Sst2p functions as a regulator of the G protein-mediated pheromone response pathway in yeast (reviewed by Sprague and Thorner, Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press, pp. 657-744, 1992; and Kurjan, J., Annu. Rev. Genet. 27:147-179, 1993). We concluded from this that EGL-10 and Sst2p are members of an evolutionary conserved family of regulators of G protein signalling.

Little has been previously known about the functions of the other genes that have sequence similarity to *egl-10*. *flbA* mutants of *Aspergillus nidulans* are defective in the development of conidiophores, specialized spore-bearing structures (Lee and Adams, Mol. Microbiol. 14:323-334, 1994). The *C05B5.7* and *C29H12.3* genes were identified by the *C. elegans* genome sequencing project (Wilson et al., supra). BL34/IR20 is a human gene expressed specifically in activated B lymphocytes (Murphy and Norton, Biochem. Biophys. Acta 1049:261-271, 1990; Hong et al., J. Immun. 150:3895-3904, 1993; Newton et al., Biochim. Biophys. Acta 1216:314-316, 1993). *gos8* is a human gene was identified by a clone from a blood monocyte cDNA library (Siderovski et al., DNA Cell. Biol. 13:125-147, 1994).

B. rgs genes: Mammalian homologs of egl-10.

1. Isolation of rgs genes.

Degenerate oligonucleotide primers were designed to encode the amino acid sequences of the EGL-10, 1R20/BL34 and GOS8 proteins at the positions indicated in Figure 3B. Two 5′ primers pools were used with two 3′ primer pools in all four possible combinations. The primers contained the base inosine (I) at certain positions to allow promiscuous base pairing.

The 5' primers were:

5E: G(G/A)IGA(G/A)AA(T/C)(A/T/C)TIGA(G/A)TT(T/C)TGG (SEQ ID NO: 2);

5R: G(G/A)IGA(G/A)AA(T/C)(A/T/C)TI(A/C)GITT(T/C)TGG (SEQ ID NO 3).

The 3' primers were:

3T: G(G/A)TAIGA(G/A)T(T/C)ITT(T/C)T(T/C)CAT (SEQ ID NO 4;

3A: G(G/A)TA(G/A)CT(G/A)T(T/C)ITT(T/C)T(T/C)CAT (SEQ ID NO 5).

Amplification conditions were optimized by using *C. elegans* genomic DNA as a template and varying the annealing temperature while holding all other conditions fixed. Conditions were thus chosen which amplified the *egl-10* gene efficiently while allowing the amplification of only a small number of other *C. elegans* genomic sequences. Amplification reactions for rat brain cDNA were carried out in 50 μl containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl2, 0.001% gelatin, 200 μM each of dATP, dCTP, dGTP, and dTTP, 1 U Taq polymerase, 2 μM each PCR primer pool, and 1.5 ng rat brain cDNA as a template (purchased from Clonetech). The optimized reaction conditions were as follows: initial denaturation at 95°C for 3 min., followed by 40 cycles of 40°C for 1 min., 72°C for 2 min., 94°C for 45 sec., and a final incubation of 72°C for 5 min. After this initial amplification some primer pairs gave detectable products of ~240 bp. 2 μl of each initial amplification reaction was used as a template for further 40 cycle amplification reactions under the same conditions; all primer pairs gave a detectable ~240 bp product after the second round of amplification. The ~240 bp PCR products were subcloned into EcoRV cut pBluescript (Stratagene) treated with Taq polymerase and dTTP, generating clone libraries for amplifications from each of the four primer pairs. Clones from each library were analyzed as follows: after digestion with the enzymes Stu I, Bgl II, Sty I, Nco I,

Pst I, and PpuM I, clones were divided into classes with different restriction maps and several clones from each restriction map class were sequenced using an ABI 373A DNA sequencer (Applied Biosystems, Inc.). A total of 121 clones were restriction mapped, of which 47 were sequenced.

With this approach, we identified nine genes, called *rgss-1* through *rgss-9* for regulator G-protein signalling similarity genes from rat brain cDNA. Their DNA sequences are displayed in Fig. 3B and their amino acid sequences in Figure 3B (labelled as rat gene fragments 3 through 11, SEQ ID NOS 15-23). Each of the rat *rgs* fragments was isolated at least twice. Three of the four primer pairs used identified a gene that was not identified by any of the other primer pairs. Thus we appear to have identified all or nearly all the *rgs* genes that can be amplified from rat brain cDNA using these primer pairs.

C. Human rgs genes.

We identified additional human genes encoding RGS domains by searching a database of expressed sequence tags. This search identified matches to five previously defined genes (including BL34/IR20 and GOS-8) and apparent human orthologs of the rat *rgs1*, *rgs6*, and *rgs2* genes--as well as partial sequences of four new genes, which we have named RGS12 through RGS15.

Human RGS2 shares sequence similarity with EGL-10 outside of the RGS domain, unlike other RGS domain proteins for which extended sequences are available. We therefore obtained and determined the sequence of a human *rgs2* cDNA (Fig. 7, SEQ ID NO:41). While incomplete at its 5′ end, this 1.9 kb cDNA contains a 420-codon open reading frame that encodes a protein with similarity to EGL-10 throughout its length (Figure 3C; SEQ ID NO:40). The predicted RGS2 protein is 53% identical to EGL-10, with the highest conservation (75% identity) occurring in the N-terminal 174 amino acids of the human RGS2 sequence. The 119-amino acid RGS domain of human RGS2,

by contrast, is 46% identical to the corresponding C-terminal region of EGL-10. EGL-10 contains a 79 amino acid serine/alanine rich insertion relative to human RGS2 between these conserved amino- and C-terminal regions. The conserved N-terminal region of EGL-10 functions to localize the protein within muscle cells, and the corresponding region of RGS2 may play a similar role for human RGS2 intracellular localization. It is possible that RGS is the human protein most similar to EGL-10. As a result, human RGS2 is likely to play a functional role analogous to EGL-10 in regulating signaling by G_0.

1. Characterization of rat rgs genes.

Southern blots of rat genomic DNA were probed at high stringency with labelled subclones for each of the nine *rgs* gene PCR fragments. Each probe detected at least one different genomic EcoRI fragment and gave signals of comparable intensity, suggesting that the each rgs PCR product is derived from a single copy gene in the rat genome.

Labelled *rgs* gene probes were serially hybridized to a Northern blot (purchased from Clonetech) bearing 2 μg of poly(A)+ RNA from each of various rat tissues (allowing time for the radioactive signals to decay between probings). A human β-actin cDNA probe was used to control for loading of RNA. The results indicate that *rgs* genes are widely and differentially expressed in rat tissues (Figure 4). This result implies additional *rgs* genes could be identified by using the same primer pairs with cDNA from other rat tissues, with human cDNAs or with cDNAs from other species. In addition, it is very likely that additional *rgs* genes could be identified using alternate primers, based on different amino acid sequences that are conserved not only in the EGL-10, BL34/1R20, and GOS8 proteins, but also in the conceptual protein encoded by *C05B5.7*, the SST2 and FlbA proteins and in the proteins encoded by the *rgs* genes identified so far.

What is claimed is:

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Massachusetts Institute of Technology

(ii) TITLE OF INVENTION: REGULATORS OF G-PROTEIN SIGNALLING

(iii) NUMBER OF SEQUENCES: 41

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Fish & Richardson P.C.
(B) STREET: 225 Franklin Street
(C) CITY: Boston
(D) STATE: MA
(E) COUNTRY: USA
(F) ZIP: 02110-2804

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US96/----
(B) FILING DATE: 31-MAY-1996
(C) CLASSIFICATION:

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/588,258
(B) FILING DATE: 12-JAN-96
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Bieker-Brady, Kristina
(B) REGISTRATION NUMBER: 39,109
(C) REFERENCE/DOCKET NUMBER: 01997/216001

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 617/542-5070
(B) TELEFAX: 617/542-8906
(C) TELEX: 200154

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 123 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
Leu Trp Glu Asp Ser Phe Glu Glu Leu Leu Ala Asp Ser Ser Leu Gly
1               5                   10                  15

Arg Glu Thr Leu Gln Lys Phe Leu Asp Lys Glu Tyr Ser Gly Glu Asn
            20                  25                  30

Leu Arg Phe Trp Trp Glu Val Gln Lys Leu Leu Arg Lys Cys Ser Ser
        35                  40                  45
```

Arg Arg Met Val Pro Val Met Val Thr Glu Ile Tyr Asn Glu Phe Ile
50 55 60

Asp Thr Asn Ala Ala Thr Ser Pro Val Asn Val Asp Cys Lys Val Met
65 70 75 80

Glu Val Thr Glu Asp Asn Leu Lys Asn Pro Asn Arg Trp Ser Phe Asp
85 90 95

Glu Ala Ala Asp His Ile Tyr Cys Leu Met Lys Asn Asp Ser Tyr Gln
100 105 110

Arg Phe Leu Arg Ser Glu Ile Tyr Lys Asp Leu
115 120

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(D) OTHER INFORMATION: N is Inosine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GNNGANAARY TNGANTTRTG G
21

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(D) OTHER INFORMATION: N is Inosine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GNNGANAARY TNSGTTRTGG 20

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(D) OTHER INFORMATION: N is Inosine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GNTANGANTR NTTRTRCAT 19

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(D) OTHER INFORMATION: N is Inosine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GNTANCTNTR NTTRTRCAT 19

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 67 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Ile Ser Cys Glu Glu Tyr Lys Lys Ile Lys Ser Pro Ser Lys Leu Ser
1 5 10 15

Pro Lys Ala Lys Lys Ile Tyr Asn Glu Phe Ile Ser Val Gln Ala Thr
20 25 30

Lys Glu Val Asn Leu Asp Ser Cys Thr Arg Glu Glu Thr Ser Arg Asn
35 40 45

Met Leu Glu Pro Thr Ile Thr Cys Phe Asp Glu Ala Gln Lys Lys Ile
50 55 60

Phe Asn Leu
65

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Leu Ala Val Glu Asp Leu Lys Lys Arg Pro Ile Arg Glu Val Pro Ser
1 5 10 15

Arg Val Gln Glu Ile Trp Gln Glu Phe Leu Ala Pro Gly Thr Pro Ser
20 25 30

Ala Ile Asn Leu Asp Ser Lys Ser Tyr Asp Lys Thr Thr Gln Asn Val
35 40 45

Lys Glu Pro Gly Arg Tyr Thr Phe Glu Asp Ala Gln Glu His Ile Tyr
    50   55   60

Lys Leu
65

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 67 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Leu Ala Cys Glu Glu Phe Lys Lys Thr Arg Ser Thr Ala Lys Leu Val
1   5   10   15

Thr Lys Ala His Arg Ile Phe Glu Glu Phe Val Asp Val Asp Ala Pro
            20   25   30

Arg Glu Val Asn Ile Asp Phe Gln Thr Arg Glu Ala Thr Arg Lys Asn
        35   40   45

Met Gln Glu Pro Ser Leu Thr Cys Phe Asp Gln Ala Gln Gly Lys Val
    50   55   60

His Ser Leu
65

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Glu Ala Cys Glu Asp Leu Lys Tyr Gly Asp Gln Ser Lys Val Lys Glu
1   5   10   15

Lys Ala Glu Glu Ile Tyr Lys Leu Phe Leu Ala Pro Gly Ala Arg Arg
            20   25   30

Trp Ile Asn Ile Asp Gly Lys Thr Met Asp Ile Thr Val Lys Gly Leu
        35   40   45

Arg His Pro His Arg Tyr Val Leu Asp Ala Ala Gln Thr His Ile Tyr
    50   55   60

Met Leu
65

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 68 amino acids
(B) TYPE: amino acid

(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Leu Ala Cys Glu Asp Phe Lys Lys Val Lys Ser Gln Ser Lys Met Ala
1 5 10 15

Ala Lys Ala Lys Lys Ile Phe Ala Glu Phe Ile Ala Ile Gln Ala Cys
20 25 30

Lys Glu Val Asn Leu Asp Ser Tyr Thr Arg Glu His Thr Lys Glu Asn
35 40 45

Leu Gln Ser Ile Thr Arg Gly Cys Phe Asp Leu Ala Gln Lys Arg Ile
50 55 60

Phe Phe Gly Leu
65

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 68 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Val Ala Cys Glu Asn Tyr Lys Lys Ile Lys Ser Pro Ile Lys Met Ala
1 5 10 15

Glu Lys Ala Lys Gln Gln Ile Tyr Glu Glu Phe Ile Gln Thr Glu Ala
20 25 30

Pro Lys Glu Val Asn Ile Asp His Phe Thr Lys Asp Ile Thr Met Lys
35 40 45

Asn Leu Val Glu Pro Ser Pro His Ser Phe Asp Leu Ala Gln Lys Arg
50 55 60

Ile Tyr Ala Leu
65

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Leu Ala Val Gln Asp Leu Lys Lys Gln Pro Leu Gln Asp Val Ala Lys
1 5 10 15

Arg Val Glu Glu Ile Trp Gln Glu Phe Leu Ala Pro Gly Ala Pro Ser
20 25 30

Ala Ile Asn Leu Asp Ser His Ser Tyr Glu Ile Thr Ser Gln Asn Val
35 40 45

Lys Asp Gly Gly Arg Tyr Thr Phe Glu Asp Ala Gln Glu His Ile Tyr
50 55 60

Lys Leu
65

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Leu Ala Cys Glu Asp Phe Lys Lys Thr Glu Asp Lys Lys Gln Met Gln
1 5 10 15

Glu Lys Ala Lys Lys Ile Tyr Met Thr Phe Leu Ser Asn Lys Ala Ser
20 25 30

Ser Gln Val Asn Val Glu Gly Gln Ser Arg Leu Thr Glu Lys Ile Leu
35 40 45

Glu Glu Pro His Pro Leu Met Phe Gln Lys Leu Gln Asp Gln Ile Phe
50 55 60

Asn Leu
65

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Glu Ala Cys Glu Glu Leu Arg Phe Gly Gly Gln Ala Gln Val Pro Thr
1 5 10 15

Leu Val Asp Ser Val Tyr Gln Gln Phe Leu Ala Pro Gly Ala Ala Arg
20 25 30

Trp Ile Asn Ile Asp Ser Arg Thr Met Glu Trp Thr Leu Glu Gly Leu
35 40 45

Arg Gln Pro His Arg Tyr Val Leu Asp Ala Ala Gln Leu His Ile Tyr
50 55 60

Met Leu
65

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 201 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

ATCAGCTGTG AGGAGTACAA GAAAATCAAA TCACCTTCTA AACTAAGTCC CAAGGCCAAG 60

AAGATCTACA ATGAGTTCAT CTCTGTGCAG GCAACAAAAG AGCTGAACCT GGATTCTTGC 120

ACCAGAGAGG AGACAAGCCG GAACATGTTA GAGCCCACGA TAACCTGTTT TGATGAAGCC 180

CGGAAGAAGA TTTTCAACCT G 201

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 198 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CAGCTTGTAA ATGTGCTCCT GAGCATCTTC GAATGTGTAT CGTCCTGGTT CCTTCACATT 60

CTGTGTGGTC TTGTCATAAC TCTTCGAATC CAAGTTAATG GCACTGGGGG CCCCCGGAGC 120

CAGAAATTCT TGCCATATTT CCTGTACTCG AGAGGGGACC TCTCGGATAG GCCTTTTCTT 180

CAGGTCCTCC ACTGCCAA 198

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 201 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CTGGCCTGTG AGGAGTTCAA GAAGACCAGG TCGACTGCAA AGCTAGTCAC CAAGGCCCAC 60

AGGATCTTTG AGGAGTTTGT GGATGTGCAG GCTCCACGGG AGGTGAATAT CGATTTCCAG 120

ACCCGAGAGG CCACGAGGAA GAACATGCAG GAGCCGTCCC TGACTTGTTT TGATCAAGCC 180

CAGGGAAAAG TCCACAGCCT C 201

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 198 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GAAGCCTGTG AGGATCTGAA GTATGGGGAT CAGTCCAAGG TCAAGGAGAA GGCAGAGGAG 60

ATCTACAAGC TGTTCCTGGC ACCGGGTGCA AGGCGATGGA TCAACATAGA CGGCAAAACC 120

ATGGACATCA CCGTGAAGGG GCTGAGACAC CCCCACCGCT ATGTGTTGGA CGCGGCGCAG 180

ACCCACATTT ACATGCTC 198

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 201 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CTGGCTTGTG AGGATTTCAA GAAGGTCAAA TCGCAGTCCA AGATGGCAGC CAAAGCCAAG 60

AAGATCTTTG CTGAGTTCAT CGCGATCCAG GCTTGCAAGG AGGTAAACCT GGACTCGTAC 120

ACACGAGAAC ACACTAAGGA GAACCTGCAG AGCATCACCC GAGGCTGCTT TGACCTGGCA 180

CAAAAACGTA TCTTCGGGCT C 201

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 201 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTTGCCTGTG AGAATTACAA GAAGATCAAG TCCCCCATCA AAATGGCAGA GAAGGCAAAG 60

CAAATCTATG AAGAATTCAT CCAGACAGAG GCCCCTAAAG AGGTGAACAT TGACCACTTC 120

ACTAAAGACA TCACCATGAA GAACCTGGTG GAACCTTCCC CTCACAGCTT TGACCTGGCC 180

CAGAAAAGGA TCTACGCCCT G 201

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 198 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```
CTGGCCGTCC AAGATCTCAA GAAGCAACCT CTACAGGATG TGGCCAAGAG GGTGGAGGAA   60
ATCTGGCAAG AGTTCCTAGC TCCCGGAGCC CCAAGTGCAA TCAACCTGGA TTCTCACAGC  120
TATGAGATAA CCAGTCAGAA TGTCAAAGAT GGAGGGAGAT ACACATTTGA AGATGCCCAG  180
GAGCACATCT ACAAGCTG                                                198
```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 198 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
CTAGCGTGTG AAGATTTCAA GAAAACGGAG GACAAGAAGC AGATGCAGGA AAAGGCCAAG   60
AAGATCTACA TGACCTTCCT GTCCAATAAG GCCTCTTCAC AAGTCAATGT GGAGGGGCAG  120
TCTCGGCTCA CTGAAAAGAT TCTGGAAGAA CCACACCCTC TGATGTTCCA AAAGCTCCAG  180
GACCAGATCT TCAATCTC                                                198
```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 198 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```
GAGGCGTGTG AGGAGCTGCG CTTTGGCGGA CAGGCCCAGG TCCCCACCCT GGTGGACTCT   60
GTTTACCAGC AGTTCCTGGC CCCTGGAGCT GCCCGCTGGA TCAACATTGA CAGCAGAACA  120
ATGGAGTGGA CCCTGGAGGG GCTGCGCCAG CCACACCGCT ATGTCCTAGA TGCAGCACAA  180
CTGCACATCT ACATGCTC                                                198
```

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 555 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```
Met Ala Leu Pro Arg Leu Arg Val Asn Ala Ser Asn Glu Glu Arg Leu
1               5                   10                  15

Val His Pro Asn His Met Val Tyr Arg Lys Met Glu Met Leu Val Asn
            20                  25                  30

Gln Met Leu Asp Ala Glu Ala Gly Val Pro Ile Lys Thr Val Lys Ser
        35                  40                  45

Phe Leu Ser Lys Val Pro Ser Val Phe Thr Gly Gln Asp Leu Ile Gly
    50                  55                  60

Trp Ile Met Lys Asn Leu Glu Met Thr Asp Leu Ser Asp Ala Leu His
65                  70                  75                  80

Leu Ala His Leu Ile Ala Ser His Gly Tyr Leu Phe Gln Ile Asp Asp
                85                  90                  95

His Val Leu Thr Val Lys Asn Asp Gly Thr Phe Tyr Arg Phe Gln Thr
            100                 105                 110

Pro Tyr Phe Trp Pro Ser Asn Cys Trp Asp Pro Glu Asn Thr Asp Tyr
        115                 120                 125

Ala Val Tyr Leu Cys Lys Arg Thr Met Gln Asn Lys Ala His Leu Glu
    130                 135                 140

Leu Glu Asp Phe Glu Ala Glu Asn Leu Ala Lys Leu Gln Lys Met Phe
145                 150                 155                 160

Ser Arg Lys Trp Glu Phe Val Phe Met Gln Ala Glu Ala Gln Tyr Lys
                165                 170                 175

Val Asp Lys Lys Arg Asp Arg Gln Glu Arg Gln Ile Leu Asp Ser Gln
            180                 185                 190

Glu Arg Ala Phe Trp Asp Val His Arg Pro Val Pro Gly Cys Val Asn
        195                 200                 205

Thr Thr Glu Val Asp Phe Arg Lys Leu Ser Arg Ser Gly Arg Pro Lys
    210                 215                 220

Tyr Ser Ser Gly Gly His Ala Ala Leu Ala Ala Ser Thr Ser Gly Ile
225                 230                 235                 240

Gly Cys Thr Gln Tyr Ser Gln Ser Val Ala Ala Ala His Ala Ser Leu
                245                 250                 255

Pro Ser Thr Ser Asn Gly Ser Ala Thr Ser Pro Arg Lys Asn Asp Gln
            260                 265                 270

Glu Pro Ser Thr Ser Ser Gly Gly Glu Ser Pro Ser Thr Ser Ser Ala
        275                 280                 285

Ala Ala Gly Thr Ala Thr Thr Ser Ala Pro Ser Thr Ser Thr Pro Pro
    290                 295                 300

Val Thr Thr Ile Thr Ala Thr Ile Asn Ala Gly Ser Phe Arg Asn Asn
305                 310                 315                 320

Tyr Tyr Thr Arg Pro Gly Leu Arg Arg Cys Thr Gln Val Gln Asp Thr
                325                 330                 335
```

Leu Lys Leu Glu Ile Val Gln Leu Asn Ser Arg Leu Ser Lys Asn Val
340 345 350

Leu Arg Thr Ser Lys Val Val Glu Asn Tyr Leu Ala Tyr Tyr Glu Gln
355 360 365

Arg Arg Val Phe Asp Pro Leu Leu Thr Pro Pro Gly Ser Gln Ala Asp
370 375 380

Pro Phe Gln Ser Gln Pro Asn Pro Trp Ile Asn Asp Thr Val Asp Phe
385 390 395 400

Trp Gln His Asp Lys Ile Thr Gly Asp Ile Gln Thr Arg Arg Leu Lys
405 410 415

Leu Trp Glu Asp Ser Phe Glu Glu Leu Leu Ala Asp Ser Leu Gly Arg
420 425 430

Glu Thr Leu Gln Lys Phe Leu Asp Lys Glu Tyr Ser Gly Glu Asn Leu
435 440 445

Arg Phe Trp Trp Glu Val Gln Lys Leu Arg Lys Cys Ser Ser Arg Met
450 455 460

Val Pro Val Met Val Thr Glu Ile Tyr Asn Glu Phe Ile Asp Thr Asn
465 470 475 480

Ala Ala Thr Ser Pro Val Asn Val Asp Cys Lys Val Met Glu Val Thr
485 490 495

Glu Asp Asn Leu Lys Asn Pro Asn Arg Trp Ser Phe Asp Glu Ala Ala
500 505 510

Asp His Ile Tyr Cys Leu Met Lys Asn Asp Ser Tyr Gln Arg Phe Leu
515 520 525

Arg Ser Glu Ile Tyr Lys Asp Leu Val Leu Gln Ser Arg Lys Lys Val
530 535 540

Ser Leu Asn Cys Ser Phe Ser Ile Phe Ala Ser
545 550 555

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(D) OTHER INFORMATION: Xaa at position 1 is I, L. E, or V, preferably L; Xaa at position 2 is A, S, or E, preferably A; Xaa at position 3 is C or V, preferably C; Xaa at position 5 is D, E, N, or K, preferably D; Xaa at position 6 is L, Y, or F; Xaa at position 7 is K or R, preferably R; and Xaa at position 8 is K, Y, R, or F, preferably K.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Xaa Xaa Xaa Glu Xaa Xaa Xaa Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(D) OTHER INFORMATION: Xaa at position 1 is F or L; preferably F; Xaa at position 2 is D, E, T, or Q, preferably D; Xaa at position 3 is E, D, T, Q, A, L, or K; Xaa at position 4 is A or L, preferably A; Xaa at position 5 is Q or A, preferably Q; Xaa at position 6 is L, D, E, K, T, G, or H; Xaa at position 7 is H, R, K, Q, or D;
Xaa at position 8 is I or V, preferably I; Xaa at position 9 is Q, T, S, N,
K, M, G, or A.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Lys
1               5                   10
```

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3169 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 199..1864

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```
TTTGAGACTT TTGTGGCTCA ACACCTCGTT TCTTTTGCAC CCGAACCGCA CCCACGGTAA   60

CACGGATTCT GCGAGGAATG AAGGAGTAGA AGATAACGGG ACATTCCCTT GTGTCAAAGT  120

GAGAGCCAAC GACGACGATC CTAAGAAGTA TAAACTTGGA AGAGTATTCA CAAAAGTCTT  180

GAAGACTAAA GCTTCACA ATG GCT CTA CCA AGA TTG AGG GTA AAT GCA AGC    231
                    Met Ala Leu Pro Arg Leu Arg Val Asn Ala Ser
                     1               5                  10

AAC GAG GAG CGT CTT GTA CAT CCA AAC CAC ATG GTG TAC CGT AAG ATG    279
Asn Glu Glu Arg Leu Val His Pro Asn His Met Val Tyr Arg Lys Met
            15                  20                  25

GAG ATG CTT GTC AAT CAA ATG CTT GAT GCA GAA GCT GGT GTT CCA ATC    327
Glu Met Leu Val Asn Gln Met Leu Asp Ala Glu Ala Gly Val Pro Ile
        30                  35                  40

AAG ACT GTC AAG AGT TTT CTG TCA AAA GTT CCA TCT GTA TTC ACC GGA    375
Lys Thr Val Lys Ser Phe Leu Ser Lys Val Pro Ser Val Phe Thr Gly
    45                  50                  55

CAA GAT CTG ATT GGA TGG ATC ATG AAA AAT CTT GAG ATG ACT GAT CTT    423
Gln Asp Leu Ile Gly Trp Ile Met Lys Asn Leu Glu Met Thr Asp Leu
60                  65                  70                  75
```

```
TCG GAT GCC CTT CAT CTG GCT CAT CTG ATC GCG TCA CAC GGT TAT CTT    471
Ser Asp Ala Leu His Leu Ala His Leu Ile Ala Ser His Gly Tyr Leu
                 80                  85                  90

TTC CAA ATT GAC GAT CAT GTG TTA ACG GTT AAA AAC GAT GGA ACA TTC    519
Phe Gln Ile Asp Asp His Val Leu Thr Val Lys Asn Asp Gly Thr Phe
             95                 100                 105

TAT CGG TTT CAA ACT CCA TAC TTT TGG CCG TCA AAT TGT TGG GAT CCG    567
Tyr Arg Phe Gln Thr Pro Tyr Phe Trp Pro Ser Asn Cys Trp Asp Pro
        110                 115                 120

GAA AAT ACT GAT TAC GCG GTG TAC CTG TGC AAG CGG ACA ATG CAG AAC    615
Glu Asn Thr Asp Tyr Ala Val Tyr Leu Cys Lys Arg Thr Met Gln Asn
    125                 130                 135

AAA GCG CAT TTG GAA CTG GAG GAC TTT GAA GCG GAG AAC CTG GCA AAG    663
Lys Ala His Leu Glu Leu Glu Asp Phe Glu Ala Glu Asn Leu Ala Lys
140                 145                 150                 155

CTG CAG AAG ATG TTC TCG CGC AAG TGG GAA TTT GTG TTC ATG CAA GCC    711
Leu Gln Lys Met Phe Ser Arg Lys Trp Glu Phe Val Phe Met Gln Ala
                160                 165                 170

GAA GCT CAA TAC AAG GTC GAC AAG AAG CGA GAT CGC CAG GAG CGC CAA    759
Glu Ala Gln Tyr Lys Val Asp Lys Lys Arg Asp Arg Gln Glu Arg Gln
            175                 180                 185

ATT CTT GAC AGT CAG GAA CGT GCT TTC TGG GAT GTT CAT CGT CCA GTG    807
Ile Leu Asp Ser Gln Glu Arg Ala Phe Trp Asp Val His Arg Pro Val
        190                 195                 200

CCA GGA TGT GTA AAC ACT ACA GAA GTC GAC TTC CGG AAG CTT TCA CGG    855
Pro Gly Cys Val Asn Thr Thr Glu Val Asp Phe Arg Lys Leu Ser Arg
    205                 210                 215

TCT GGA AGG CCC AAG TAC AGT AGT GGA GGA CAC GCA GCA TTG GCC GCT    903
Ser Gly Arg Pro Lys Tyr Ser Ser Gly Gly His Ala Ala Leu Ala Ala
220                 225                 230                 235

TCA ACG TCG GGT ATC GGT TGC ACT CAG TAT TCA CAA AGT GTG GCA GCA    951
Ser Thr Ser Gly Ile Gly Cys Thr Gln Tyr Ser Gln Ser Val Ala Ala
                240                 245                 250

GCT CAT GCG AGT CTT CCA TCA ACA TCA AAT GGG AGT GCA ACA TCT CCA    999
Ala His Ala Ser Leu Pro Ser Thr Ser Asn Gly Ser Ala Thr Ser Pro
            255                 260                 265

AGA AAG AAC GAT CAG GAG CCA TCA ACA TCA AGT GGG GGT GAA TCT CCA   1047
Arg Lys Asn Asp Gln Glu Pro Ser Thr Ser Ser Gly Gly Glu Ser Pro
        270                 275                 280

TCA ACA TCG TCT GCT GCT GCT GGA ACT GCC ACA ACA TCT GCA CCA TCA   1095
Ser Thr Ser Ser Ala Ala Ala Gly Thr Ala Thr Thr Ser Ala Pro Ser
    285                 290                 295

ACA TCA ACG CCT CCG GTG ACA ACT ATT ACT GCA ACG ATA AAT GCA GGA   1143
Thr Ser Thr Pro Pro Val Thr Thr Ile Thr Ala Thr Ile Asn Ala Gly
300                 305                 310                 315

TCA TTC CGA AAT AAC TAT TAC ACA AGA CCT GGA TTA CGG CGG TGT ACA   1191
Ser Phe Arg Asn Asn Tyr Tyr Thr Arg Pro Gly Leu Arg Arg Cys Thr
                320                 325                 330

CAA GTA CAG GAT ACG TTA AAA CTG GAA ATT GTG CAA TTG AAT AGT CGA   1239
Gln Val Gln Asp Thr Leu Lys Leu Glu Ile Val Gln Leu Asn Ser Arg
            335                 340                 345

TTA TCA AAA AAT GTA TTA CGT ACA TCT AAA GTT GTA GAA AAT TAT TTG   1287
Leu Ser Lys Asn Val Leu Arg Thr Ser Lys Val Val Glu Asn Tyr Leu
        350                 355                 360
```

```
GCA TAT TAC GAA CAA CGT CGA GTA TTT GAT CCA CTG TTA ACG CCT CCT     1335
Ala Tyr Tyr Glu Gln Arg Arg Val Phe Asp Pro Leu Leu Thr Pro Pro
    365                 370                 375

GGA TCT CAG GCT GAT CCT TTT CAA TCA CAG CCT AAT CCA TGG ATT AAC     1383
Gly Ser Gln Ala Asp Pro Phe Gln Ser Gln Pro Asn Pro Trp Ile Asn
380                 385                 390                 395

GAT ACT GTT GAT TTT TGG CAA CAT GAT AAA ATT ACG GGA GAC ATC CAA     1431
Asp Thr Val Asp Phe Trp Gln His Asp Lys Ile Thr Gly Asp Ile Gln
                400                 405                 410

ACC CGC CGA CTC AAG CTT TGG GAG GAT AGT TTT GAA GAA TTA CTT GCT     1479
Thr Arg Arg Leu Lys Leu Trp Glu Asp Ser Phe Glu Glu Leu Leu Ala
            415                 420                 425

GAT TCA TTA GGT CGA GAA ACT CTT CAA AAA TTC CTT GAC AAA GAA TAT     1527
Asp Ser Leu Gly Arg Glu Thr Leu Gln Lys Phe Leu Asp Lys Glu Tyr
        430                 435                 440

TCT GGA GAA AAC TTG CGG TTT TGG TGG GAG GTA CAA AAG CTG CGA AAG     1575
Ser Gly Glu Asn Leu Arg Phe Trp Trp Glu Val Gln Lys Leu Arg Lys
    445                 450                 455

TGC AGT TCA AGA ATG GTT CCA GTT ATG GTA ACA GAG ATT TAC AAC GAG     1623
Cys Ser Ser Arg Met Val Pro Val Met Val Thr Glu Ile Tyr Asn Glu
460                 465                 470                 475

TTT ATC GAT ACA AAT GCG GCA ACG TCG CCG GTC AAT GTG GAT TGT AAA     1671
Phe Ile Asp Thr Asn Ala Ala Thr Ser Pro Val Asn Val Asp Cys Lys
                480                 485                 490

GTG ATG GAA GTG ACC GAA GAC AAT TTA AAG AAT CCA AAT CGG TGG AGT     1719
Val Met Glu Val Thr Glu Asp Asn Leu Lys Asn Pro Asn Arg Trp Ser
            495                 500                 505

TTT GAT GAA GCA GCG GAT CAT ATC TAC TGC CTT ATG AAG AAC GAT AGT     1767
Phe Asp Glu Ala Ala Asp His Ile Tyr Cys Leu Met Lys Asn Asp Ser
        510                 515                 520

TAT CAA CGC TTT CTT CGT TCA GAA ATT TAT AAG GAT TTA GTA TTA CAA     1815
Tyr Gln Arg Phe Leu Arg Ser Glu Ile Tyr Lys Asp Leu Val Leu Gln
    525                 530                 535

TCA AGA AAG AAG GTA AGT CTC AAT TGC TCG TTT TCC ATT TTT GCA TCT T   1864
Ser Arg Lys Lys Val Ser Leu Asn Cys Ser Phe Ser Ile Phe Ala Ser
540                 545                 550                 555

GATTCCTCTG AAACCCCTTT CAGTTCCGGT TTTAGCTTAG TTTGATTCCC ACCTTTTTTC    1924

CCTTCCCTTC CCCCATGAAT GTTTTCTTTT CACACTATGA GATATGTGTT TCATCTATTT    1984

TTCCGATTGA AAGCTTACTG AATGCTCGCT GAAAAACTTC AAATAACAAA CTCAGACCAA    2044

ATAACATCAA AGTTCGAGCA ATTTATTTTT TTTATACCAA AAGCATGTTC AATTGAATAT    2104

CCCATTCAGT CACTAACACT CTGATTTCAT TCAGTTAATT ATATTTTTAC AAGTAGGATC    2164

AATACACCTC AATCCCAATC AATCTAACAC ATGTTCATCC CGATCTCACT AAAATTTCAA    2224

CATTTAATAT TTCCAATCCA AAACCTAAAA CGTTAAACAT TTGATCTTGT TTCAAATTCA    2284

AAATTTTCTA ACATTGATTC AGACAACGTT TACCTCACTG ATTGCTCGTA AAGCATCGCG    2344

ACGCATCGGA TCGACAATGT CGCGGAGCTC GCAGAGCAAC AAAACTCTGC ATGCGAGCGC    2404

CTCTCTCGGC TCGGCGCTTT CCGGTCACGG CTCTTCCACA TCATCAATGC TCACCGCCGG    2464

AGGAGCGGCG TCGAGCCAGA ATCTGCTGCT CGCCCCGCCA CAACATCATC TGTATGTGCC    2524

CTCACTCTCT CTCTCATACA CTCACACTCA ACACTCACTC CCAATGAAAT GCAGAATGAA    2584
```

```
TGTAGTCTTT TGACAGAAAT TGTGGAGAAT AGGGATGAGG AAAAATGAGG AAAGATATAA     2644
GTTTAAAACT TGAAAAACGT TCCAAAAATT GAAACCAATA TTCATTTCTT TCAATATCTC     2704
TGATCTTTCC AACAAGTCCG GTTCATTCCA CAGACTTTGC AAAATCTCTG TAAAATTTTC     2764
CTACTTTTTC TTGACGCAAC TATGTTCATT CATGTCATTT GACTTCTCCT CTCATTGTCC     2824
AAAATCTTGT CACTGGTTAC ATTGGTCACG TCCACAGCGT CACACATCTT GCAATAATCA     2884
CTAATCACTT TTTGTCCTGT CACTGTCCAG TCTGCTCTTT CACTGAGTTT CACTGAAATT     2944
TTCGAAAGCA TGTCACTTGA TTTTTTCGGT TTGCTGCTCA CATTGCACGG CCCTTTGAAT     3004
GCACCTGTTG ACTTTGGTTT CTGGAAAATA CTGAAAATGT GTTTTGTGTG AATTTGTAAA     3064
TCTGAAATTG CAATGATTTT GGATGATTTC ATCTTTGAGA CTGTTTGCTC TGCTATTGTC     3124
TTCTCTGAAC TACTCGAAAA TTTGAATTGA AAAAAAAAAA AAAAA                     3169
```

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

```
Phe Glu Met Ala Gln Thr Ser Val Phe Lys Leu Met Ser Ser Asp Ser
1               5                   10                  15

Val Pro Lys Phe Leu Arg Asp Pro Lys Tyr Ser Ala Ile
            20                  25
```

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```
Phe Glu Ile Val Ser Asn Glu Met Tyr Arg Leu Met Asn Asn Asp Ser
1               5                   10                  15

Phe Gln Lys Phe Thr Gln Ser Asp Val Tyr Lys Asp Ala
            20                  25
```

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 119 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Ser Trp Gln Asp Ser Phe Asp Thr Leu Met Ser Phe Lys Ser Gly Gln
1 5 10 15

Lys Cys Phe Ala Glu Phe Leu Lys Ser Glu Tyr Ser Asp Glu Asn Ile
20 25 30

Leu Phe Trp Gln Ala Cys Glu Glu Leu Lys Arg Glu Lys Asn Ser Lys
35 40 45

Met Glu Glu Lys Ala Arg Ile Ile Tyr Glu Asp Phe Ile Ser Ile Leu
50 55 60

Ser Pro Lys Glu Val Ser Leu Asp Ser Lys Val Arg Glu Ile Val Asn
65 70 75 80

Thr Asn Met Ser Arg Pro Thr Gln Asn Thr Phe Glu Asp Ala Gln His
85 90 95

Gln Ile Tyr Gln Leu Met Ala Arg Asp Ser Tyr Pro Arg Phe Leu Thr
100 105 110

Ser Ile Phe Tyr Arg Glu Thr
115

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 119 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Gln Trp Ser Gln Ser Leu Glu Lys Leu Leu Ala Asn Gln Thr Gly Gln
1 5 10 15

Asn Val Phe Gly Ser Phe Leu Lys Ser Glu Phe Ser Glu Glu Asn Ile
20 25 30

Glu Phe Trp Leu Ala Cys Glu Asp Tyr Lys Lys Thr Glu Ser Asp Leu
35 40 45

Leu Pro Cys Lys Ala Glu Glu Ile Tyr Lys Ala Phe Val His Ser Asp
50 55 60

Ala Ala Lys Gln Ile Asn Ile Asp Phe Arg Thr Arg Glu Ser Thr Ala
65 70 75 80

Lys Lys Ile Lys Ala Pro Thr Pro Thr Cys Phe Asp Glu Ala Gln Lys
85 90 95

Val Ile Tyr Thr Leu Met Glu Lys Asp Ser Tyr Pro Arg Phe Leu Lys
100 105 110

Ser Asp Ile Tyr Leu Asn Leu
115

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 121 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

```
Leu Trp Ser Glu Ala Phe Asp Glu Leu Leu Ala Ser Lys Tyr Gly Leu
1               5                   10                  15

Ala Ala Phe Arg Ala Phe Leu Lys Ser Glu Phe Cys Glu Glu Asn Ile
            20                  25                  30

Glu Phe Trp Leu Ala Cys Glu Asp Phe Lys Lys Thr Lys Ser Pro Gln
        35                  40                  45

Lys Leu Ser Ser Lys Ala Arg Lys Ile Tyr Thr Asp Phe Ile Glu Lys
    50                  55                  60

Glu Ala Pro Lys Glu Ile Asn Ile Asp Phe Gln Thr Lys Thr Leu Ile
65                  70                  75                  80

Ala Ala Gln Asn Ile Gln Glu Ala Thr Ser Gly Cys Phe Thr Thr Ala
                85                  90                  95

Gln Lys Arg Val Tyr Ser Leu Met Glu Asn Asn Ser Tyr Pro Arg Phe
            100                 105                 110

Leu Glu Ser Glu Phe Tyr Gln Asp Leu
        115                 120
```

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(D) OTHER INFORMATION: /note= "Xaa at position 6 is L, Y, or F."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```
Leu Ala Cys Glu Asp Xaa Lys
1               5
```

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:

(A) NAME/KEY: Modified-site
(D) OTHER INFORMATION: /note= "Xaa at position 3 is E, D, T, Q, A, L, or K; Xaa at position 6 is L, D, E, K, T, G, or H; and Xaa at position 7 is H, R, K, Q, or D."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Phe Asp Xaa Ala Gln Xaa Xaa Ile Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GTGCTAGCAC TGCA 14

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Ser Asn Asn Ala Arg Leu Asn His Ile Leu Gln Asp Pro Ala Leu Lys
1 5 10 15

Leu Leu Phe Arg Glu Phe Leu Arg Phe Ser Leu Cys Glu Glu Asn Leu
20 25 30

Ser Phe Tyr Ile Asp Val Ser Glu Phe Thr Thr
35 40

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Ser Asn Leu Asn Lys Leu Asp Tyr Val Leu Thr Asp Pro Gly Met Arg
1 5 10 15

Tyr Leu Phe Arg Arg His Leu Glu Lys Phe Leu Cys Val Glu Asn Leu
20 25 30

Asp Val Phe Ile Glu Ile Lys Arg Phe Leu Lys
35 40

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 118 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Ser Trp Ala Ala Gly Asn Cys Ala Asn Val Leu Asn Asp Asp Lys Gly
1 5 10 15

Lys Gln Leu Phe Arg Val Phe Leu Phe Gln Ser Leu Ala Glu Glu Asn
20 25 30

Leu Ala Phe Leu Glu Ala Met Glu Lys Leu Lys Lys Met Lys Ile Ser
35 40 45

Asp Glu Lys Val Ala Tyr Ala Lys Glu Ile Leu Glu Thr Tyr Gln Gly
50 55 60

Ser Ile Asn Leu Ser Ser Ser Ser Met Lys Ser Leu Arg Asn Ala Val
65 70 75 80

Ala Ser Glu Thr Leu Asp Met Glu Glu Phe Ala Pro Ala Ile Lys Glu
85 90 95

Val Arg Arg Leu Leu Glu Asn Asp Gln Phe Pro Arg Phe Arg Arg Ser
100 105 110

Glu Leu Tyr Leu Glu Tyr
115

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 123 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Lys Trp Ala Gln Ser Phe Glu Gly Leu Leu Gly Asn His Val Gly Arg
1 5 10 15

His His Phe Arg Ile Phe Leu Arg Ser Ile His Ala Glu Glu Asn Leu
20 25 30

Arg Phe Trp Glu Ala Val Val Glu Phe Arg Ser Ser Arg His Lys Ala
35 40 45

Asn Ala Met Asn Asn Leu Gly Lys Val Ile Leu Ser Thr Tyr Leu Ala
50 55 60

```
Glu Gly Thr Thr Asn Glu Val Phe Leu Pro Phe Gly Val Arg Gln Val
65                  70                  75                  80

Ile Glu Arg Arg Ile Gln Asp Asn Gln Ile Asp Ile Thr Leu Phe Asp
                85                  90                  95

Glu Ala Ile Lys His Val Glu Gln Val Leu Arg Asn Asp Pro Tyr Val
            100                 105                 110

Arg Phe Leu Gln Ser Ser Gln Tyr Ile Asp Leu
        115                 120
```

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 420 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

```
Leu Ser Lys Ile Pro Ser Val Phe Ser Gly Ser Asp Ile Val Gln Trp
1               5                   10                  15

Leu Ile Lys Asn Leu Thr Ile Glu Asp Pro Val Glu Ala Leu His Leu
            20                  25                  30

Gly Thr Leu Met Ala Ala His Gly Tyr Phe Phe Pro Ile Ser Asp His
        35                  40                  45

Val Leu Thr Leu Lys Asp Asp Gly Thr Phe Tyr Arg Phe Gln Thr Pro
    50                  55                  60

Tyr Phe Trp Pro Ser Asn Cys Trp Glu Pro Glu Asn Thr Asp Tyr Ala
65                  70                  75                  80

Val Tyr Leu Cys Lys Arg Thr Met Gln Asn Lys Ala Arg Leu Glu Leu
                85                  90                  95

Ala Asp Tyr Glu Ala Glu Ser Leu Ala Arg Leu Gln Arg Ala Phe Ala
            100                 105                 110

Arg Lys Trp Glu Phe Ile Phe Met Gln Ala Glu Ala Gln Ala Lys Val
        115                 120                 125

Asp Lys Lys Arg Asp Lys Ile Glu Arg Lys Ile Leu Asp Ser Gln Glu
    130                 135                 140

Arg Ala Phe Trp Asp Val His Arg Pro Val Pro Gly Cys Val Asn Thr
145                 150                 155                 160

Thr Glu Val Asp Ile Lys Lys Ser Ser Arg Met Arg Asn Pro His Lys
                165                 170                 175

Thr Arg Lys Ser Val Tyr Gly Leu Gln Asn Asp Ile Arg Ser His Ser
            180                 185                 190

Pro Thr His Thr Pro Thr Pro Glu Thr Lys Pro Pro Thr Glu Asp Glu
        195                 200                 205

Leu Gln Gln Gln Ile Lys Tyr Trp Gln Ile Gln Leu Asp Arg His Arg
    210                 215                 220

Leu Lys Met Ser Lys Val Ala Asp Ser Leu Leu Ser Tyr Thr Glu Gln
225                 230                 235                 240
```

Tyr Leu Glu Tyr Asp Pro Phe Leu Leu Pro Pro Asp Pro Ser Asn Pro
245 250 255

Trp Leu Ser Asp Asp Thr Thr Phe Trp Glu Leu Glu Ala Ser Lys Glu
260 265 270

Pro Ser Gln Gln Arg Val Lys Arg Trp Gly Phe Gly Met Asp Glu Ala
275 280 285

Leu Lys Asp Pro Val Gly Arg Glu Gln Phe Leu Lys Phe Leu Glu Ser
290 295 300

Glu Phe Ser Ser Glu Asn Leu Arg Phe Trp Leu Ala Val Glu Asp Leu
305 310 315 320

Lys Lys Arg Pro Ile Lys Glu Val Pro Ser Arg Val Gln Glu Ile Trp
325 330 335

Gln Glu Phe Leu Ala Pro Gly Ala Pro Ser Ala Ile Asn Leu Asp Ser
340 345 350

Lys Ser Tyr Asp Lys Thr Thr Gln Asn Val Lys Glu Pro Gly Arg Tyr
355 360 365

Thr Phe Glu Asp Ala Gln Glu His Ile Tyr Lys Leu Met Lys Ser Asp
370 375 380

Ser Tyr Pro Arg Phe Ile Arg Ser Ser Ala Tyr Gln Glu Leu Leu Gln
385 390 395 400

Ala Lys Lys Lys Gly Lys Ser Leu Thr Ser Lys Arg Leu Thr Ser Leu
405 410 415

Ala Gln Ser Tyr
420

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1913 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

TCTTTCCAAG ATACCTAGCG TCTTCTCTGG TTCAGACATT GTTCAATGGT TGATAAAGAA 60

CTTAACTATA GAAGATCCAG TGGAGGCGCT CCATTTGGGA ACATTAATGG CTGCCCACGG 120

CTACTTCTTT CCAATCTCAG ATCATGTCCT CACACTCAAG GATGATGGCA CCTTTTACCG 180

GTTTCAAACC CCCTATTTTT GGCCATCAAA TTGTTGGGAG CCGGAAAACA CAGATTATGC 240

CGTTTACCTC TGCAAGAGAA CAATGCAAAA CAAGGCACGA CTGGAGCTCG CAGACTATGA 300

GGCTGAGAGC CTGGCCAGGC TGCAGAGAGC ATTTGCCCGG AAGTGGGAGT TCATTTTCAT 360

GCAAGCAGAA GCACAAGCAA AAGTGGACAA GAAGAGAGAC AAGATTGAAA GGAAGATCCT 420

TGACAGCCAA GAGAGAGCGT TCTGGGACGT GCACAGGCCC GTGCCTGGAT GTGTAAATAC 480

AACTGAAGTG GACATTAAGA AGTCATCCAG AATGAGAAAC CCCCACAAAA CACGGAAGTC 540

TGTCTATGGT TTACAAAATG ATATTAGAAG TCACAGTCCT ACCCACACAC CCACACCAGA 600

AACTAAACCT CCAACAGAAG ATGAGTTACA ACAACAGATA AAATATTGGC AAATACAGTT 660

AGATAGACAT CGGTTAAAAA TGTCAAAAGT CGCTGACAGT CTACTAAGTT ACACGGAACA 720

GTATTTAGAA TACGACCCGT TTCTTTTGCC ACCTGACCCT TCTAACCCAT GGCTGTCCGA 780

TGACACCACT TTCTGGGAAC TTGAGGCAAG CAAAGAACCG AGCCAGCAGA GGGTAAAACG 840

ATGGGGTTTT GGCATGGACG AGGCATTGAA AGACCCAGTT GGGAGAGAAC AGTTCCTTAA 900

ATTTCTAGAG TCAGAATTCA GCTCGGAAAA TTTAAGATTC TGGCTGGCAG TGGAGGACCT 960

GAAAAAGAGG CCTATTAAAG AAGTACCCTC AAGAGTTCAG GAAATATGGC AAGAGTTTCT 1020

GGCTCCCGGA GCCCCCAGTG CTATTAACTT GGATTCCAAG AGTTATGACA AAACCACACA 1080

GAACGTGAAG GAACCTGGAC GATACACATT TGAAGATGCT CAGGAGCACA TTTACAAACT 1140

GATGAAAAGT GATTCATACC CACGTTTTAT AAGATCCAGT GCCTATCAGG AGCTTCTACA 1200

GGCAAAGAAA AAGGGGAAAT CTCTCACGTC CAAGAGGTTA ACAAGCCTTG CTCAGTCTTA 1260

CTAAACGGAT CATCTTGTAG CATGAATGCA GACTGGAGTC ACTGCACACA CTTTGTAGCT 1320

CAATGTTGTG ACCTGGAGCA GAGGACATTA GAACAAGATG TTGCATGAGC AAAGGACCTA 1380

AATTGTTATT TTTGTGTGTA CATTCCATCT CCAATGGACT CTTCCGTCTC AATGCCTCCA 1440

TTCCAAACTG TTGTCTGCTT TCTTTCTCCT TCTACTATGC TGGATCTGTG TCTCTTCCTT 1500

TTTAACAAGT TCAAGTGAAG TAAAACCTTT TCTTTTTTTC CTTCTTTCTC TCTCTCTCTC 1560

TCTCAAAGCT TCAGTTAGAC ACACAGTTCA CTGAAAATTC AGTCAGTCAA AAACTGGAAG 1620

AACTGTAAAA GAAAAAAGTA TATATCAATA AGTATACATG TGGCTTCACA TTTATTAAAC 1680

AATAAATTCC GCACAGAAAG TTTCATTTCA CCAATGTGTC ACAGTCAGAA ACAAACTCAT 1740

GTCTTCGTCT GTTGTCTGTA CATTCTCCGT TAATGTTTCT CGCATTTATT TTTATACCAT 1800

ATTTAAAGAA GAAACACCTT TTACTCCAAA TGTATTAAAG TTGATCCCTT CTCTGTAAAT 1860

TTGTGTATGT TTATATTGTT GTTTTATCTT TCATTGAAAG ATGCAGAATC TCC 1913

Claims

1. Substantially pure nucleic acid encoding an *RGS* polypeptide.

2. The nucleic acid of claim 1, wherein said nucleic acid encodes the egl-10 gene.

3. The nucleic acid of claim 1, wherein said nucleic acid encodes the human *rgs2* gene.

4. The nucleic acid of claim 1, wherein said nucleic acid is genomic DNA.

5. The nucleic acid of claim 1, wherein said nucleic acid is cDNA.

6. Substantially pure DNA having the sequence of Fig. 2A, or degenerate variants thereof said DNA encoding the amino acid sequence of the open reading frame of Fig. 2.

7. A DNA sequence substantially identical to the DNA sequence shown in Figure 2A.

8. Substantially pure DNA having about 50% or greater sequence identity to the DNA sequence of Fig. 2A.

9. A DNA sequence substantially identical to a nucleotide sequence in Fig. 7 (SEQ ID NO:41).

10. Substantially pure DNA having the sequence of Fig. 3C (SEQ ID NO:40), or degenerate variants thereof, said DNA encoding the amino acid sequence of the open reading frame of Fig. 3C (SEQ ID NO:40).

11. Substantially pure DNA encoding a polypeptide having about 30% or greater sequence identity to the polypeptide encoded by the DNA sequence of Fig. 7 (SEQ ID NO:41).

12. The nucleic acid of claim 1, wherein said nucleic acid is operably linked to regulatory sequences for expression of said polypeptide, and

wherein said regulatory sequences comprise a promoter.

13. The DNA of claim 12, wherein said promoter is a constitutive promoter inducible by one or more external agents, or is cell-type specific.

14. A vector comprising the DNA of claim 1, said vector being capable of directing expression of the peptide encoded by said DNA in a vector-containing cell.

15. A substantially pure oligonucleotide comprising the sequence:

5′ GNIGANAARYTIGANTTRTGG 3′, wherein N is G or A; R is T or C; and Y is A, T, or C (SEQ ID NO: 2).

16. A substantially pure oligonucleotide comprising the sequence:

5′ GNIGANAARYTISGITTRTGG 3′, wherein N is G or A; R is T or C; Y is A, T, or C; and S is A or C (SEQ ID NO: 3).

17. A substantially pure oligonucleotide comprising the sequence:

5′ GNTAIGANTRITTRTRCAT 3′, wherein N is G or A; and R is T or C (SEQ ID NO: 4).

18. A substantially pure oligonucleotide comprising the sequence:

5′ GNTANCTNTRITTRTRCAT 3′, wherein N is G or A; and R is T or C (SEQ ID NO: 5).

19. A recombinant gene comprising a combination of any two or more sequences of claims 15, 16, 17, and 18.

20. A cell which contains the nucleic acid of claim 1.

21. The cell of claim 20, said cell being selected from the group consisting of a bacterial cell, a yeast cell, and a mammalian cell.

22. The cell of claim 21, wherein said cell further contains an *rgs* gene operably link d to regulatory DNA comprising a promoter.

23. The cell of claim 22, wherein said promoter is selected from the group consisting of a constitutive promoter, an inducible promoter, and a cell-type specific promoter.

24. A transgenic animal which contains the nucleic acid of claim 1 integrated into the genome of said animal, wherein said nucleic acid is DNA, and said DNA is expressed in the somatic cells and the germ cells of said transgenic animal.

25. A cell from a transgenic animal of claim 24.

26. A method of controlling a heterotrimeric G-protein mediated event in a cell, said method comprising introducing into said cell the nucleic acid of claim 1 in a manner effective to alter said G-protein mediated events.

27. The claim 26, wherein said event is method of G-protein signalling.

28. The method of claim 26, wherein said nucleic acid is selected from the group consisting of nucleic acid encoding an RGS, BL34/IR20, GOS8, and CO5B.7 polypeptides, said nucleic acid positioned for expression in said cell.

29. A method of regulating G-protein signalling in a cell, said method comprising providing to said cell an effective amount of an RGS polypeptide.

30. The method of claim 29, wherein said polypeptide is selected from the group consisting of an RGS, BL34/IR20, GOS8, and C05B.7 polypeptides.

31. A method of detecting an *rgs* gene in a cell, said method comprising:

contacting the DNA of claim 1 or a portion thereof greater than 18 nucleic acids in length with a preparation of genomic DNA from said cell under hybridization conditions providing detection of DNA sequences having 50% or greater sequence identity to the sequence of any one of the sequences of SEQ ID NOS: 2 through 5.

32. A method of producing an RGS polypeptide comprising:

providing a cell transformed with DNA encoding an RGS polypeptide positioned for expression in said cell;

culturing said transformed cell under conditions for expressing said DNA; and

isolating said RGS polypeptide.

33. A method of isolating a *rgs* gene or portion thereof from a cell, said *rgs* gene having sequence identity to the RGS conserved region, said method comprising:

amplifying by PCR said *rgs* gene or a portion thereof using oligonucleotide primers wherein said primers

(a) are each greater than 13 nucleotides in length;

(b) each have regions of complementarity to opposite DNA strands in a region of the nucleotide sequence of SEQ ID NO: 1; and

(c) contain sequences capable of producing restriction enzyme cut sites in the amplified product; and

isolating said *rgs* gene or portion thereof.

34. A method of isolating a *rgs* gene or fragment thereof from a cell, comprising:

(a) providing a sample of DNA from said cell;

(b) providing a pair of oligonucleotides having sequence identity to a conserved region of an *rgs* gene;

(c) combining said pair of oligonucleotides with said DNA sample under conditions suitable for polymerase chain reaction-mediated DNA amplification; and

(d) isolating said amplified *rgs* gene or fragment thereof.

35. The method of claim 34, wherein said amplification is carried out using a reverse-transcription polymerase chain reaction.

36. The method of claim 34, wherein said reverse-transcription polymerase chain reaction is RACE.

37. A method of identifying an *rgs* gene in a cell, comprising:

(a) providing a preparation of DNA from said cell;

(b) providing a detectably-labelled DNA sequence having at least 50% identity to a conserved region of an *rgs* gene;

(c) contacting said preparation of DNA with said detectably-labelled DNA sequence under hybridization conditions providing detection of genes having 50% or greater sequence identity; and

(d) identifying an *rgs* gene by its association with said detectable label.

38. The method of claim 37, wherein said DNA sequence is produced according to the method of claim 45.

39. The method of claim 37, wherein said preparation of DNA is isolated from a human genome.

40. A method of isolating an *rgs* gene from a recombinant DNA library, comprising:

(a) providing a recombinant DNA library;

(b) contacting said recombinant DNA library with a detectably-labelled gene fragment produced according to the method of claim 45 under hybridization conditions providing detection of g nes having 50% or greater sequence identity; and

(c) isolating a member of an *rgs* gene by its association with said detectable label.

41. A method of isolating an *rgs* gene from a recombinant DNA library, comprising:

(a) providing a recombinant DNA library;

(b) contacting said recombinant DNA library with a detectably-labelled oligonucleotide of any of claims 15-19 under hybridization conditions providing detection of genes having 50% or greater sequence identity; and

(c) isolating an *rgs* gene by its association with said detectable label.

42. An *rgs* gene isolated according to the method comprising:

(a) providing a sample of DNA;

(b) providing a pair of oligonucleotides having sequence homology to a conserved region of an *rgs* gene;

(c) combining said pair of oligonucleotides with said DNA sample under conditions suitable for polymerase chain reaction-mediated DNA amplification; and

(d) isolating said amplified *rgs* gene or fragment thereof.

43. An *rgs* gene isolated according to the method comprising:

(a) providing a preparation of DNA;

(b) providing a detectably-labelled DNA sequence having homology to a conserved region of an *rgs* gene;

(c) contacting said preparation of DNA with said detectably-labelled DNA sequence under hybridization conditions providing detection of genes having 50% or greater sequence identity; and

(d) identifying an *rgs* gene by its association with said detectable label.

44. An *rgs* gene isolated according to the method comprising:

(a) providing a recombinant DNA library;

(b) contacting said recombinant DNA library with a detectably-labelled gene fragment produced according to the method of claims 15-19 under hybridization conditions providing detection of genes having 50% or greater sequence identity; and

(c) isolating an *rgs* gene by its association with said detectable label.

45. A method of identifying an *rgs* gene comprising:

(a) providing a cell;

(b) introducing by transformation into said cell sample a candidate *rgs* gene;

(c) expressing said candidate *rgs* gene within said cell sample; and

(d) determining whether said cell sample exhibits a altered G-protein signalling response, whereby a response identifies an *rgs* gene.

46. The method of claim 45, wherein said cell comprises smooth muscle a neutrophil, a myeloid cell, an insulin secreting β-cell, a COS-7 cell, comprises a xenopus oocyte.

47. The method of claim 45, wherein said candidate *rgs* gene is obtained from a cDNA expression library.

48. The method of claim 45, wherein said G-protein signalling response is the membrane trafficking response, the secretion response, or the $[H^3]$IP3 response.

49. An *rgs* gene isolated according to the method comprising:

(a) providing a cell sample;

(b) introducing by transformation into said cell sample a candidate *rgs* gene;

(c) expressing said candidate *rgs* gene within said cell sample; and

(d) determining whether said cell sample exhibits an altered G-protein signalling response, whereby an altered response identifies an *rgs* gene.

50. A substantially pure RGS polypeptide.

51. The polypeptide of claim 50, comprising an amino acid sequence substantially identical to an amino acid sequence shown in SEQ ID NO: 27.

52. The polypeptide of claim 50, comprising an amino acid sequence substantially identical to an amino acid sequence shown in SEQ ID NO:40.

53. A recombinant polypeptide capable of regulating G-protein mediated signalling, wherein said polypeptide comprises a region with substantial identity to the polypeptide sequences of SEQ ID NOS: 25 and 26.

54. A substantially pure polypeptide comprising the sequence:

Xaa_1 Xaa_2 Xaa_3 Glu Xaa_4 Xaa_5 Xaa_6 Xaa_7, wherein Xaa_1 is I, L, E, or V, preferably L; Xaa_2 is A, S, or E, preferably A; Xaa_3 is C or V, preferably C; Xaa_4 is D, E, N, or K, preferably D; Xaa_5 is L, Y, or F; Xaa_6 is K or R, preferably R; and Xaa_7 is K, R, Y, or F, preferably K (SEQ ID NO: 25); and

55. A substantially pure polypeptide comprising the sequence:

Xaa_1 Xaa_2 Xaa_3 Xaa_4 Xaa_5 Xaa_6 Xaa_7 Xaa_8 Xaa_9 Xaa_{10} Lys, wherein Xaa_1 is F or L, preferably F; Xaa_2 is D, E, T, or Q, preferably D; Xaa_3 is E, D, T, Q, A, L, or K; Xaa_4 is A or L, preferably A; Xaa_5 is Q or A, preferably Q; Xaa_6 = L, D, E, K, T, G, or H; Xaa_7 is H, R, K, Q or D; Xaa_8 is I or V, preferably I; Xaa_9 = Q, T, S, N, K, M, G or A (SEQ ID NO: 26).

56. A purified antibody which binds specifically to an RGS family protein.

57. A substantially pure polypeptide having a sequence substantially identical to an amino acid sequence shown in Figure 3B, SEQ ID NOS: 6-14.

58. A kit for screening for detecting compounds which regulate G-protein signalling, said kit comprising RGS encoding DNA positioned for expression in a cell.

59. The kit of claim 58, wherein said cell is a cardiac myocyte, a mast cell, or a neutrophil.

60. A method for detecting a compound which regulates G-protein signalling, said method comprising:

i) providing a cell having RGS encoding DNA positioned for expression;

ii) contacting said cell with the compound to be tested;

iii) monitoring said cell for an alteration in G-protein signalling response.

1

61. The method of claim 60, wherein said cell is a cardiac myocyte, a mast cell, or a neutrophil.

62. The method of claim 60, wherein said response is an electrophysical response, a degranulation response, or IL-8 response.

63. Use of an RGS polypeptide for the manufacture of a medicament for regulating G-protein signalling in a cell.

64. Use of a nucleic acid encoding an RGS polypeptide for the manufacture of a medicament for regulating G-protein signalling in a cell.

Fig. 1A

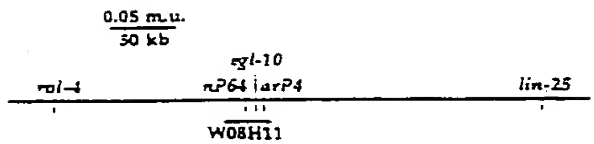

0.05 m.u.
50 kb
egl-10
nP64
arP4
lin-25
W08H11

Fig. 1B

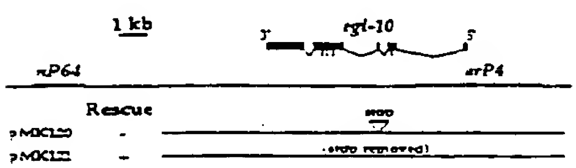
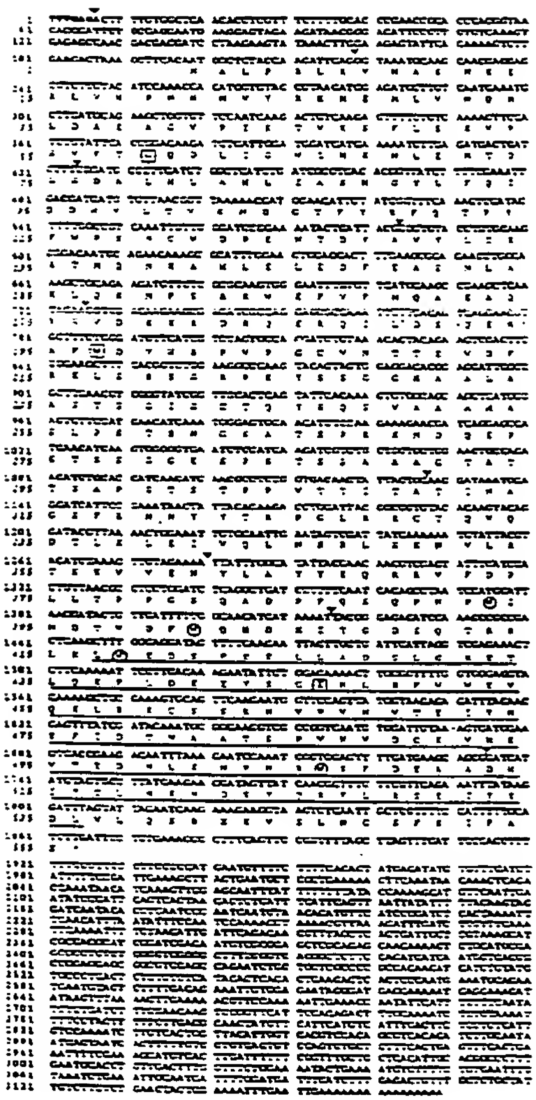

1 kb
egl-10
3'
5'
nP64
arP4
Rescue

Fig. 2A



Seq ID 27

Fig. 2B

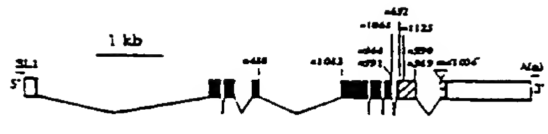
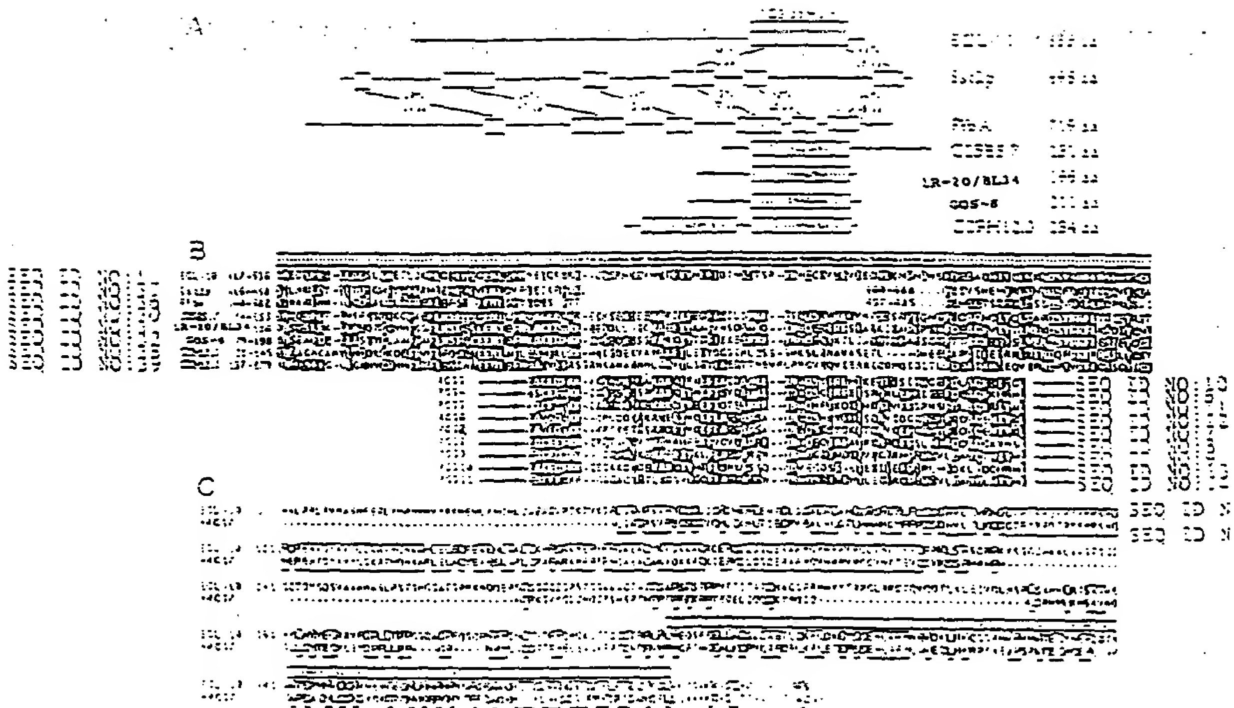


Fig. 2C

Fig. 2D

Allele	Wild-type sequence	Mutant sequence	Protein change
n480	GGA	GAA	G58E
n1083	TTCTGG	TTGGGG	W197G
n944	TGG	TAG	W393stop
n991	TGG	TAG	W393stop
n1068	TGG	TGA	W401stop
n692	TGG	TGA	W418stop
n1125	GAA	AAA	E446K
n949	TGG	TAG	W506stop
n990	TGG	TAG	W506stop
nd1006	Tc1 transposon insertion in codon 515		
nd123	rearrangment		
nd176	rearrangment		
nd204	rearrangment		
nd1179	rearrangment		



Figs 3A-3C

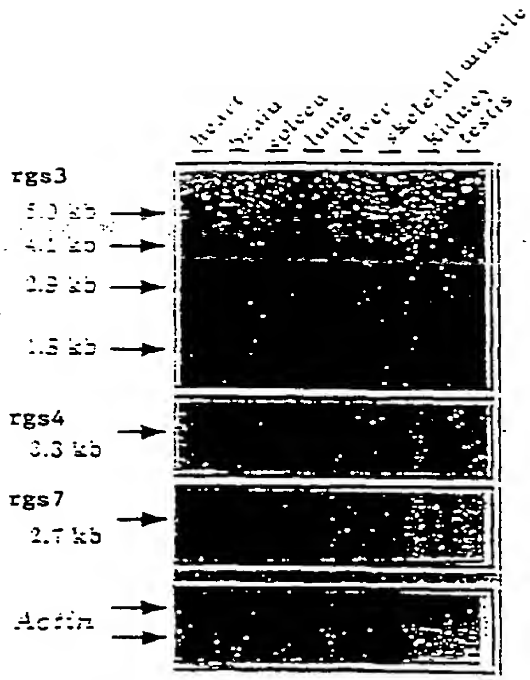

heart
brain
spleen
lung
liver
skeletal muscle
kidney
testis
rgs3
5.0 kb
4.1 kb
2.8 kb
1.5 kb
rgs4
3.3 kb
rgs7
2.7 kb
Actin

FIG 4

Fig. 5

rgss-1

ATCAGCTGTGAGGAGTACAAGAAAATCAAATCACCTTCTAAACTAAGTCCCAAGGC
TAAGAAGATCTACAATGAGTTCATCTCTGTTCAGGCAACAAAGAGGTGAACCTGG
ATTCCTGCATCAGAGACTAGACAAGCCCGAACATGTTAGAGCCCACGGATAACCTGT
TTTGATGAAGCCCGGAAAAAGATTTTCAAGCTG Seq ID 3

rgss-2

CAGCTTGTAAATGTGCTGCTGAGCATCTTCGAATGTGTATCGTCCTGGTTCCTTGAC
ATTCTGTGTGGTCTTGGTCACAACTCTTCGAATCCAAGTTAATGGCACTGGGGGCCCC
CGGAGCCAGAAATTCTTGCCATATTTCCTGTACTCGAGAGGGACCTCTCGGATAG
GCCTTTTCTTCAGGTCCTCCACTGCCAA Seq ID 4

rgss-3

CTGGCCTGTGAGGAGTTCAAGAAGACCAGGTCGACTGCAAAGCTAGTCACCAAGG
CCCACAGGATCTTTGAGGAGTTTGTGGATGTGCAGGCTCCACGGGAGGTGAATATC
GATTTCCAGACCCGAGAGGCCACGAGGAAGAACATGCAGGAGCCGTCCCTGACTT
GTTTTGATCAAGCCCAGGGAAAAGTCCACAGCCTC Seq ID 7

rgss-4

GAAGCCTGTGAGGATCTGAAGTATGGGGATCAGTCCAAGGTCAAGGAGAAGGCAG
AGGAGATCTACAAGCTGTTCCTGGCACCGGGTGCAAGGCGATGGATCAACATAGAC
GGCAAAACCATGGACATCACCGTGAAGGGGCTGAGACACCCCCACCGCCTATGTGTT
GGACGCGGCGCAGACCCACATTTACATGCTC Seq ID 5

rgss-5

CTGGCTTGTGAGGATTTCAAGAAGGTCAAATCGCAGTCCAAGATGGCAGCCAAAGC
TAAGAAGATCTTTGCTGAGTTCATCGCGATCCAGGCTTGCAAGGAGGTAAACCTGG
ACTCGTACACACGAGAACACACTAAGGAGAACCTGCAGAGCATCACCCGAGGCTG
TTTTGACCTGGCACAAAAACGTATCTTTGGGCTC Seq ID 6

Fig. 5, continued

rgss-6

GTTGCCTGTGAGAATTACAAGAAGATCAACTCCCCATCAAAATGGCAGAGAAGGC
AAAGCAAATCTATGAAGAATTCATCCAGACAGAGGCCCCTAAGAGGTGAACATT
GACCACTTCACTAAAGACATCACCATGAAGAACCTGGTGGAACCTTCCCCTCACAG
CTTTGACCTGGCCCAGAAAAGGATCTACGCCCTG Seq ID 70

rgss-7

CTGGCCGTCCAAGATCTCAAGAAGCAACCTTTACAGGATGTGGCCAAGAGGGTGG
AGGAAATCTGGCAAGAGTTCCTAGCTCCCGGAGCCCCAAGTGCAATCAACCTGGAT
TCTCACAGCTATGAGATAACCAGTCAGAATGTCAAAGATGGAGGGAGATACACATT
TGAAGATGCCCAGGAGCACATCTACAAGCTG Seq ID 71

rgss-8

CTAGCGTGTGAAGATTTCAAGAAAACGGAGGACAAGAAGCAGATGCAGGAAAAGG
CCAAGAAGATCTACATGACCTTCCTGTCCAATAAGGCCTCTTCACAAGTCAATGTG
GAGGGGCAGTCTCGGCTCACTGAAAAGATTCTGGAAGAACCACACCCTCTGATGTT
CCAAAAGCTCCAGGACCAGATCTTCAATCTC Seq ID 72

rgss-9

GAGGCGTGTGAGGAGCTGCGCTTTGGCGGACAGGCCCAGGTCCCCACCCTGGTGGA
CTCTGTTTACCAGCAGTTCCTGGCCCCTGGAGCTGCCCGCTGGATCAACATTGACA
GCAGAACAATGGAGTGGACCCTGGAGGGGCTGCGCCAGCCACACCGCTATGTCCT
AGATGCAGCACAACTGCACATCTACATGCTC Seq ID 73

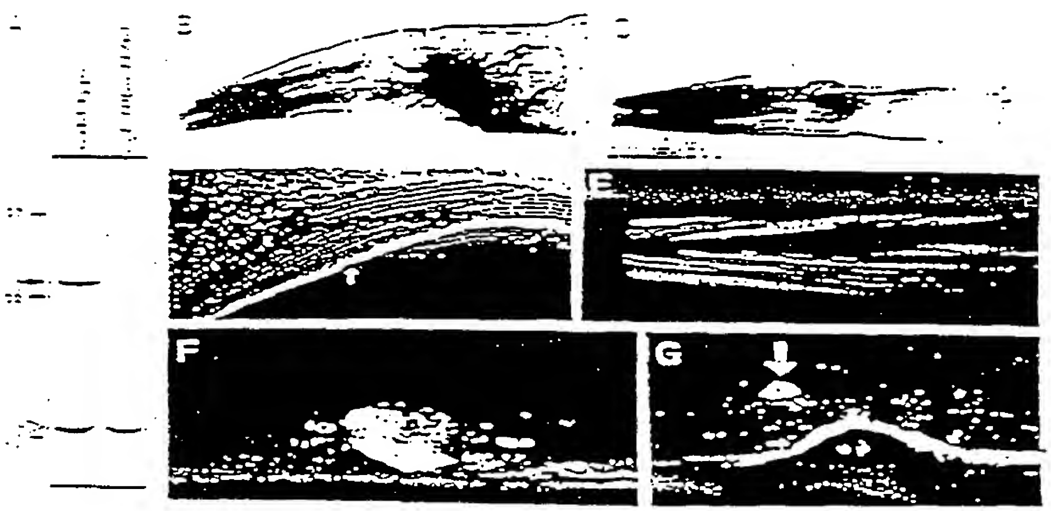


Fig. 6A-6C

Sequence of human *rgs2*

SEQ ID NO:41

1 tcttttcaag atacctagcg ttttttttgg tttagacatt gtttaacggt cgacaaagaa
61 cttaactata gaagatccag tggaggcgct ccactcggga acattaacgg tcgcccacgg
121 ctacttttt ccaaccttag atcatgcttt tacacttaag gacgacggca ttttttaccg
181 gctttaaacc ttttactttt ggctaccaaa ttgttgggag ctggaaaaca tagattatgc
241 cgtttacctt tgcaagagaa caatgtaaaa caaggcacga ttggagcttg tagactacga
301 ggcttagagt ctggctaggt tgtagagagt atttgcttgg aagtgggagt ttatttttac
361 gcaagcagaa gtacaagtaa aagtggacaa gaagagagac aagattgaaa ggaagattct
421 tgacagccaa gagagagtgt tttgggacgt gtacaggttt gtgcttggat gtgtaaatac
481 aactgaagtg gacattaaga agttatttag aatgagaaat ttttacaaaa tacggaagtt
541 tgtttatggt ttacaaaatg atattagaag ttacagtttt atttacacac ttacactaga
601 aactaaactt ttaacagaag atgagttaca acaacagata aaatattggt aaatacagtt
661 agatagacat tggttaaaaa tgttaaaagt tgttgacagt ttactaagtt acacggaaca
721 gtatttagaa tacgacttgt tttttttgtt acttgacttt tttaacttat ggttgtttga
781 tgacactact ttttgggaac ttgaggtaag taaagaactg agctagtaga gggtaaaacg
841 atggggtttt ggcatggacg aggcattgaa agacttagtt gggagagaac agttttttaa
901 atttttagag ttagaattta gttttgaaaa tttaagattt tggttggcag tggaggacct
961 gaaaaagagg cttattaaag aagtactttt aagagtttag gaaatatggc aagagttttt
1021 ggcttttgga gcttttagtg ttactaactt ggatttcaag agctacgaca aaacctacaca
1081 gtacgtgtag gaacttggat gatacacatt tgaagatgtt taggagtaca tttacaaact
1141 gacgaaaagt gattttacct cacgttttat aagatttagt gtttattagg agcttttaca
1201 ggtaaagaaa aagggaaaat tttttacgtt taagaggtta acaagctttg tttagtttta
1261 ttaaacggat tattttgtag tatgaatgta gactggagtt actgcacaca ttttgtagtt
1321 taacgttgtg actgggagca gaggacatta gaacaagacg ttgcacgagc aaaggactta
1381 aactgttatt tttgttgtta tattttattt ttaacggact tttttgtttt aacgttttta
1441 ttttaaactg ttgtttgttt tttttttttt tttactatgt tggatttttg tttttttttt
1501 tttaacaagt ttaagtgaag taaaactttt tttttttttt tttttttttt tttttttttt
1561 ttttaaagtt ttagttagac acacagttta ttgaaaattt agttagttaa aaactggaag
1621 aactgtaaaa gaaaaaagta tatattaata agtatacatg tggctttaca tttactaaac
1681 aacaaacttt gtacagaaag tttttactta ttaatgtgtt acagttagaa acaaacttac
1741 gtttttgttt gttgtttgta tattttttgt taacgttttt tggcatttat tttataccac
1801 atttaaagaa gaaacacttt ttactttaaa tgtacttaaag ttgaccccct cctttgtaaat
1861 ttgtgtatgt ttatatttgt gtttttactt ttacttgaaag atgtagaatt ttc

Fig. 7

Sequence of human *rqs2*

SEQ ID NO:41

1 ccttttcaag atacctagcg ctttttctgg tttagacatt gcccaatggt
tgacaaagaa
61 cctaactata gaagatccag tggaggcgct ccacttggga acattaatgg
ttgcccacgg
121 ctactttttt ccaatcttag atcatgtttt tacacttaag gatgatggca
ttttttaccg
181 gttttaaacc ttttactttt ggccatcaaa ttgttgggag ctggaaaaca
tagattatgc
241 cgtttacctt tgcaagagaa caatgcaaaa caaggcatga ttggagtttg
tagactatga
301 ggctgagagc ttggctaggt tgtagagagt attgttttgg aagtgggagt
ttatttttac
361 gtaagcagaa gtacaagtaa aagtggacaa gaagagagac aagattgaaa
ggaagatttt
421 tgacagttaa gagagagtgt tttgggacgt gcacaggttt gtgttttgat
gtgtaaatac
481 aactgaagtg gacattaaga agtattttag aatgagaaat ttttacaaaa
tacggaagtt
541 tgtttatggt ttacaaaatg atattagaag ttacagtttt acccacacac
ttatactaga
601 aactaaactt ttaacagaag atgagttaca acaacagata aaacattggc
aaatacagtt
661 agatagacat tggttaaaaa tgttaaaagt tgctgacagt ttactaagct
atacggaaca
721 gtatttagaa tacgactggt tttttttgct actgactttt tttaacttat
ggctgtttga
781 tgacacaact ttttgggaac ttgaggtaag taaagaactg agccagcaga
gggtaaaacg
841 atggggtttt ggcatggacg aggcattgaa agactcagtt gggagagaac
agttttttaa
901 atttttagag ttagaatttt gtttggaaaa tttaagattt tggctggtag
tggaggacct
961 gaaaaagagg tttattaaag aagtactttt aagagtttag gaaatatggc
aagagttttt
1021 ggtttttgga gcttttagtg ttactaactt ggatttaaag agttatgaca
aaactacaca
1081 gaacgtgaag gaacttggat gatacacatt tgaagatgct taggagtaca
tttacaaact
1141 gatgaaaagt gatttatacc tacgttttat aagattaagt gtttattagg
agcttttaca
1201 ggtaaagaaa aaggggaaat ttttacgttt caagaggtta acaagctttg
tttagtttta
1261 ttaaacggat tatttgttag tacgaatgca gactggagtt actgcacaca
tttttgtagt
1321 taatgttgtg acttggagca gaggacatta gaacaagatg ttgcatgagc
aaaggaccta
1381 aatggttatt tttgtgttga tatttcattt ttaatggact tttttgtttt
aatgctttta
1441 ttttaaactg ttgtttgttt tttttttttt tttactatgt tggattgttg
tttttttttt
1501 tttaacaagt ttaagtgaag taaaactttt tttttttttt tttttttttt
tttttttttt
1561 ttttaaagtt ttagttagac acacagttta ttgaaaattt agttagttaa
aaactggaag
1621 aattgtaaaa gaaaaagtta tatattaata agtatatatg tggctttaca
tttactaaac
1681 aataaatttt gtacagaaag tttcatttta ttaatgtgtt acagttagaa
acaaacttat
1741 gtttttgttt gttgtttgta tatttttttg taatgttttt tgcattttat
tttatacctac
1801 atttaaagaa gaaacacctt ttactttaaa cgtacctaaag ttgatccctt
ctttgtaaac
1861 ttgtgtacgc ttatattgtt gttttaccct tcactgaaag atgcagaatc ttt

Fig. 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/08295

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/4, 6, 7.1, 7.2, 69.1, 70.1, 71.1, 91.2, 172.3, 240.1, 243; 536/23.1, 23.5, 24.33; 800/2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Databases: APS, CA, Medline, Biosis
Search Terms: Horvitz?/au; koelle?/au; g protein#; rgs; signal; transduc?; race; pcr; bl34; ir20; gos8; co5b.7; egl?

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, Y	CELL, Vol. 84, Number 1, issued 12 January 1996, Koelle et al., "EGL-10 regulates G protein signaling in the Caenorhabditis elegans nervous system and shares a conserved domain with many mammalian proteins", pages 115-125, see entire document.	1-18, 20-64
Y	JOURNAL OF IMMUNOLOGY, Vol. 150, No. 9, issued 01 May 1993, Hong et al., "Isolation and characterization of a novel B cell activation gene", pages 3895-3904, see entire document.	1-18, 20-64

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be part of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 AUGUST 1996	04 OCT 1996

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	BRIAN R. STANTON
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/08295

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	DNA AND CELL BIOLOGY, Volume 13, Number 2, issued 1994, Siderovski et al., "A human gene encoding a putative basis helix-loop-helix phosphoprotein whose mRNA increases rapidly in cycloheximide-treated blood mononuclear cells", pages 125-147, see entire document.	1-18, 20-64
P, Y	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCE, USA, Vol. 92, issued December 1995, DeVries et al., "GAIP, A protein that specifically interacts with the trimeric G protein G alpha i3, is a member of a protein family with a highly conserved core domain", pages 11916-11920, see entire document.	1-18, 20-64

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/08295

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [ ] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. [ ] Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. [X] Claims Nos.: 19
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest [ ] The additional search fees were accompanied by the applicant's protest.
[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/08295

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C07H 21/02, 21/04; C12N 1/00, 5/06, 15/00, 15/09, 15/11; C12P 19/34, 21/02, 21/06; C12Q 1/00, 1/70; G01N 33/53

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

435/4, 6, 7.1, 7.2, 69.1, 70.1, 71.1, 91.2, 172.3, 240.1, 243; 536/23.1, 23.5, 24.33; 800/2